(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

# CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 November 2001 (29.11.2001)

PCT

(10) International Publication Number
## WO 01/090148 A2

(51) International Patent Classification[7]: C07K 14/00

(21) International Application Number: PCT/US01/16283

(22) International Filing Date: 17 May 2001 (17.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
| | | |
|---|---|---|
| 60/205,518 | 19 May 2000 (19.05.2000) | US |
| 60/213,956 | 22 June 2000 (22.06.2000) | US |
| 60/215,105 | 28 June 2000 (28.06.2000) | US |
| 60/218,947 | 14 July 2000 (14.07.2000) | US |
| 60/221,488 | 27 July 2000 (27.07.2000) | US |

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): SANJANWALA, Madhu, S. [US/US]; 210 Sylvia Court, Los Altos, CA 94024 (US). WALIA, Narinder, K. [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). TRIBOULEY, Catherine, M. [FR/US]; 1121 Tennessee Street #5, San Francisco, CA 94107 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). GANDHI, Ameena, R. [US/US]; 837 Roble Avenue #1, Menlo Park, CA 94025 (US). DING, Li [CN/US]; 3353 Alma Street #146, Palo Alto, CA 94306 (US). YAO, Monique, G. [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). LAL, Preeti [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). HAFALIA, April [US/US]; 2227 Calle de Primavera, Santa Clara, CA

95054 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). RANKUMAR, Jaya [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— without international search report and to be republished upon receipt of that report

(48) Date of publication of this corrected version:
4 July 2002

(15) Information about Correction:
see PCT Gazette No. 27/2002 of 4 July 2002, Section II

For two-letter codes and other abbreviations. refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NEUROTRANSMITTER TRANSPORTERS

(57) Abstract: The invention provides human neurotransmitter transporter (NTT) and polynucleotides which identify and encode NTT. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of NTT.

## NEUROTRANSMITTER TRANSPORTERS

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of neurotransmitter transporters

5    and to the use of these sequences in the diagnosis, treatment, and prevention of transport, neurological

and psychiatric disorders, and in the assessment of the effects of exogenous compounds on the

expression of nucleic acid and amino acid sequences of neurotransmitter transporters.

### BACKGROUND OF THE INVENTION

10    The human nervous system, which regulates all bodily functions, is composed of the central

nervous system (CNS), consisting of the brain and spinal cord, and the peripheral nervous system

(PNS), consisting of afferent neural pathways for conducting nerve impulses from sensory organs to the

CNS, and efferent neural pathways for conducting motor impulses from the CNS to effector organs.

The PNS can be further divided into the somatic nervous system, which regulates voluntary motor

15    activity such as for skeletal muscle, and the autonomic nervous system, which regulates involuntary

motor activity for internal organs such as the heart, lungs, and viscera.

The central nervous system (CNS) is composed of more than 100 billion neurons at the spinal

cord level, the lower brain level, and the higher brain or cortical level. Neurons transmit electric or

chemical signals between cells. The spinal cord, a thin, tubular extension of the central nervous

20    system within the bony spinal canal, contains ascending sensory and descending motor pathways, and

is covered by membranes continuous with those of the brainstem and cerebral hemispheres. The

spinal cord contains almost the entire motor output and sensory input systems of the trunk and limbs,

and neuronal circuits in the cord also control rhythmic movements, such as walking, and a variety of

reflexes. The lower areas of the brain such as the medulla, pons, mesencephalon, cerebellum, basal

25    ganglia, substantia nigra, hypothalamus, and thalamus control unconscious activities including arterial

pressure and respiration, equilibrium, and feeding reflexes, such as salivation. Emotions, such as

anger, excitement, sexual response, and reaction to pain or pleasure, originate in the lower brain. The

cerebral cortex or higher brain is the largest structure, consisting of a right and a left hemisphere

interconnected by the corpus callosum. The cerebral cortex is involved in sensory, motor, and

30    integrative functions related to perception, voluntary musculoskeletal movements, and the broad

range of activities associated with consciousness, language, emotions, and memory. The cerebrum

functions in association with the lower centers of the nervous system.

A nerve cell (neuron) contains four regions, the cell body, axon, dendrites, and axon terminal.

The cell body contains the nucleus and other organelles. The dendrites are processes which extend

outward from the cell body and receive signals from sense organs or from the axons of other neurons. These signals are converted to electrical impulses and transmitted to the cell body. The axon, whose size can range from one millimeter to more than one meter, is a single process that conducts the nerve impulse away from the cell body. Cytoskeletal fibers, including microtubules and neurofilaments, run

5    the length of the axon and function in transporting proteins, membrane vesicles, and other macromolecules from the cell body along the axon to the axon terminal. Some axons are surrounded by a myelin sheath made up of membranes from either an oligodendrocyte cell (CNS) or a Schwann cell (PNS). Myelinated axons conduct electrical impulses faster than unmyelinated ones of the same diameter. The axon terminal is at the tip of the axon away from the cell body. (See Lodish, H. et al.

10   (1986) Molecular Cell Biology Scientific American Books New York NY, pp. 715-719.)

CNS-associated proteins have roles in neuronal signaling, cell adhesion, nerve regeneration, axon guidance, neurogenesis, and other functions. Certain CNS-associated proteins form an integral part of a membrane or are attached to a membrane. For example, neural membrane protein 35 (NMP35) is closely associated with neuronal membranes and is known to be highly expressed in the rat

15   adult nervous system. (Schweitzer, B. et al. (1998) Mol. Cell. Neurosci. 11:260-273.) Synaptophysin (SY) is a major integral membrane protein of small synaptic vesicles. The chromosomal location of SY in human and mouse is on the X chromosome in subbands Xp11.22-p11.23. This region has been implicated in several inherited diseases including Wiskott-Aldrich syndrome, three forms of X-linked hypercalciuric nephrolithiaisis, and the eye disorders retinitis pigmentosa 2, congenital stationary night

20   blindness, and Aland Island eye disease. (Fisher, S. E. et al. (1997) Genomics 45:340-347.) Peripherin or retinal degeneration slow protein (rds) is an integral membrane glycoprotein that is present in the rims of photoreceptor outer segment disks. In mammals, rds is thought to stabilize the disk rim through heterophilic interactions with related nonglycosylated proteins. Rds is a mouse neurological mutation that is characterized by abnormal development of rod and cone photoreceptors followed by their slow

25   degeneration. (Kedzierski, W.J. et al. (1999) Neurochem. 72:430-438.)

43 KD postsynaptic protein or acteylcholine receptor-associated 43 KD protein (RAPSYN) is thought to play a role in anchoring or stabilizing the nicotinic acetylcholine receptor at synaptic sites. RAPSYN is involved in membrane association and may link the nicotinic acetylcholine receptor to the underlying postsynaptic cytoskeleton. (Buckel, A. et al. (1996) Genomics 35:613-616.) Neuritin is a

30   protein whose gene is known to be induced by neural activity and by neurotrophins which promotes neuritogenesis. Neuraxin is a structural protein of the rat central nervous system that is believed to be immunologically related to microtubule-associated protein 5 (MAP5). Neuraxin is a novel type of neuron-specific protein which is characterized by an unusual amino acid composition, 12 central heptadecarepeats and putative protein and membrane interaction sites. The gene encoding neuraxin is

unique in the haploid rat genome and is conserved in higher vertebrates. Neuraxin is implicated in neuronal membrane-microtubule interactions and is expressed throughout the rodent central nervous system (CNS). (Rienitz, A. et al. (1989) EMBO J. 8:2879-2888.)

NudC, a nuclear movement protein, interacts with the lissencephaly gene product Lis1, a
5     protein involved in neuronal migration. People with Miller-Dieker syndrome (MDS) or isolated lissencephaly sequence (ILS) have a hemizygous deletion or mutation in the LIS1 gene. Both conditions are characterized by a smooth cerebral surface, a thickened cortex with four abnormal layers, and misplaced neurons. LIS1 is highly expressed in the ventricular zone and the cortical plate. The interaction of Lis1 with NudC, in conjunction with the MDS and ILS phenotypes, raises the possibility
10    that nuclear movement in the ventricular zone is closely related to neuronal fates and to cortical architecture. (Morris, S. M. et al. (1998) Curr. Biol. 8:603-606.)

CNS-associated proteins can also be phosphoproteins. For example, ARPP-21 (cyclic AMP-regulated phosphoprotein) is a cytosolic neuronal phosphoprotein that is highly enriched in the striatum and in other dopaminoceptive regions of the brain. The steady-state level of ARPP-21 mRNA
15    is developmentally regulated. But, in the neonatal and mature animal, ARPP-21 mRNA is not altered following 6-hydroxydopamine lesions of the substantia nigra or by pharmacologic treatments that upregulate the D1- or D2-dopamine receptors. (Ehrlich, M. E. et al. (1991) Neurochem. 57:1985-1991.)

CNS-associated signaling proteins may contain PDZ domains. PDZ domains have been found
20    in proteins which act as adaptors in the assembly of multifunctional protein complexes involved in signaling events at surfaces of cell membranes. PDZ domains are generally found in membrane-associated proteins including neuronal nitric oxide synthase (NOS) and several dystrophin-associated proteins. (Ponting, C. P. et al. (1997) Bioessays 19:469-479.)

CNS-associated proteins may also contain epidermal growth factor (EGF) domains. The Notch
25    proteins are transmembrane proteins which contain extracellular regions of repeated EGF domains. Notch proteins, such as the Drosophila melanogaster neurogenic protein Notch, are generally involved in the inhibition of developmental processes. Other members of the Notch family are the lin-12 and glp-1 genes of Caenorhabditis elegans. Genetic studies indicate that the lin-12 and glp-1 proteins act as receptors in specific developmental cell interactions which may be involved in certain embyronic
30    defects. (Tax, F. E. et al. (1994) Nature 368:150-154.) Pecanex, a maternal-effect neurogenic locus of D. melanogaster is believed to encode a large transmembrane protein. In the absence of maternal expression of the pecanex gene, an embryo develops severe hyperneuralization similar to that characteristic of Notch mutant embryos. (LaBonne, S. G. et al. (1989) Dev. Biol. 136:1-116.) Other CNS-associated signaling proteins contain WW domains. The WW domain is a protein motif with two

highly conserved tryptophans. It is present in a number of signaling and regulatory proteins, including Huntingtin interacting protein. Several fibroblast growth factor (FGF) homologous factors (*i.e.*, FHF polypeptides) have also been implicated in nervous system development based on mRNA expression patterns in mouse and human tissues. Members of the FHF family of polypeptides are structurally

5 distinct from prototypic FGFs, consistent with the unusual role of these FGF-related proteins (Smallwood, P.M. et al. (1996) Proc. Natl. Acad. Sci. U.S.A. 93:9850-9857 and Hartung, H. et al. (1997) Mech. Dev. 64:31-39).

Alzheimer's disease (AD) is a degenerative disorder of the CNS which causes progressive memory loss and cognitive decline during mid to late adult life. AD is characterized by a wide range of

10 neuropathologic features including amyloid deposits and intra-neuronal neurofibrillary tangles. Although the pathogenic pathway leading to neurodegeneration and AD is not well understood, at least three genetic loci that confer genetic susceptibility to the disease have been identified. (Schellenberg, G.D. (1995) Proc. Natl. Acad. Sci. 92:8552-8559; Sherrington, R. et al. (1995) Nature 375:754-760.)

Neuronal Thread Proteins (NTP) are a group of immunologically related molecules found in

15 the brain and neuroectodermal tumor cell lines. NTP expression is increased in neuronal cells during proliferation, differentiation, brain development, in Alzheimer's disease (AD) brains, and in pathological states associated with regenerative nerve sprouting (de la Monte, S.M. et al. (1996) J. Neuropathol. Exp. Neurol. 55:1038-1050). Monoclonal antibodies generated to a recombinant NTP, AD7c-NTP, isolated from an end-stage AD brain library, showed high levels of NTP

20 immunoreactivity in perikarya, neuropil fibers, and white matter fibers of AD brain tissue. In vitro studies also demonstrated NTP upregulation, phosphorylation, and translocation from the perikarya to cell processes and growth cones during growth factor-induced neuitic sprouting and neuronal differentiation. Additionally, increased NTP immunoreactivity was found in Down syndrome brains beginning in the second decade, prior to establishment of widespread AD neurodegeneration, and at

25 an age when a low-level or an absence of NTP expression was observed in control brains. These findings indicated that abnormal expression and accumulation of NTP in brain may be an early marker of AD neurodegeneration in Down syndrome (de la Monte, S.M. et al. (1996) J. Neurol. Sci. 135:118-125). Furthermore, the increased expression and accumulation of NTP in AD brain tissue was paralleled by corresponding elevations of NTP in cerebrospinal fluid (CSF), and elevated levels

30 of NTP were detectable in the CSF early in the course of the disease.

Astrocytomas, and the more malignant glioblastomas, are the most common primary tumors of the brain, accounting for over 65% of primary brain tumors. These tumors arise in glial cells of the astrocyte lineage. Following infection by pathogens, astrocytes function as antigen-presenting cells and modulate the activity of lymphocytes and macrophages. Astrocytomas constitutively express many

4

cytokines and interleukins that are normally produced only after infection by a pathogen (de Micco, C. (1989) J. Neuroimmunol. 25:93-108). In the course of identifying genes related to astrocyte differentiation, one cDNA was isolated from an astrocytoma cDNA library that encodes a protein structurally related to the plant pathogenesis-related (PR) proteins (Murphy, E.V. et al. (1995) Gene

5    159:131-135). The glioma pathogenesis-related protein (GliPR) is highly expressed in glioblastoma, but not in fetal or adult brain, or in other nervous system tumors. PR proteins are a family of small (10-20 kDa), protease resistant proteins induced in plants by viral infections, such as tobacco mosaic virus. The synthesis of PR proteins is believed to be part of a primitive immunological response in plants (van Loon, L.C. (1985) Plant Mol. Biol. 4:111-116). GliPR shares up to 50% homology with

10   the PR-1 protein family over a region that comprises almost two thirds of the protein, including a conserved triad of amino acids, His-Glu-His, appropriately spaced to form a metal-binding domain (Murphy et al., supra).

Fe65-like protein (Fe65L2), a new member of the Fe65 protein family, is one of the ligands that interacts with the cytoplasmic domain of Alzheimer beta-amyloid precursor protein (APP). Transgenic

15   mice expressing APP are known to simulate some of the prominent behavioral and pathological features of Alzheimer's disease, including age-related impairment in learning and memory, neuronal loss, gliosis, neuritic changes, amyloid deposition, and abnormal tau phosphorylation. Proteins that interact with the cytoplasmic domain of APP provide new insights into the physiological function of APP and, in turn, into the pathogenesis of Alzheimer's disease. (Duilio, A. et al. (1998) Biochem. J. 330:513-519.)

20   Contact from one neuron to another occurs at a specialized site called the synapse. At this site, the axon terminal from one neuron (the presynaptic cell) sends a signal to another neuron (the postsynaptic cell). Synapses may be connected either electrically or chemically. An electrical synapse consists of gap junctions connecting the two neurons, allowing electrical impulses to pass directly from the presynaptic to the postsynaptic cell. In a chemical synapse, the axon terminal of the presynaptic cell

25   contains membrane vesicles containing a particular neurotransmitter molecule. A change in electrical potential at the nerve terminal results in the influx of calcium ions through voltage-gated channels which triggers the release of the neurotransmitter from the synaptic vesicle by exocytosis. The neurotransmitter rapidly diffuses across the synaptic cleft separating the presynaptic nerve cell from the postsynaptic cell. The neurotransmitter then binds receptors and opens transmitter-gated ion

30   channels located in the plasma membrane of the postsynaptic cell, provoking a change in the cell's electrical potential. This change in membrane potential of the postsynaptic cell may serve either to excite or inhibit further transmission of the nerve impulse. Presynaptic calcium channel activity is modulated by cysteine-string proteins (CSPs). CSPs are secretory vesicle proteins that function in neurotransmission as well as in exocytosis in other cell-types. The effect of CSPs on calcium levels is

likely to be downstream of calcium release and is likely to involve exocytosis, possibly in connection with G-proteins (Magga, J.M. et al. (2000) Neuron 28:195-204; and Dawson-Scully, K. et al. (2000) J. Neurosci. 20:6039-6047; and Chamberlain, L.H. et al. (2001) J. Cell Sci. 114:445-455).

Neurotransmitters comprise a diverse group of some 30 small molecules which include
5    acetylcholine, monoamines such as serotonin, dopamine, and histamine, and amino acids such as gamma-aminobutyric acid (GABA), glutamate, and aspartate, and neuropeptides such as endorphins and enkephalins. (McCance, K.L. and Huether, S.E. (1994) PATHOPHYSIOLOGY, The Biologic Basis for Disease in Adults and Children, 2nd edition, Mosby, St. Louis, MO, pp 403-404.) Many of these molecules have more than one function and the effects may be excitatory, e.g. to depolarize the
10   postsynaptic cell plasma membrane and stimulate nerve impulse transmission, or inhibitory, e.g. to hyperpolarize the plasma membrane and inhibit nerve impulse transmission.

Neurotransmitters and their receptors are targets of pharmacological agents aimed at controlling neurological function. For example GABA is the major inhibitory neurotransmitter in the CNS, and GABA receptors are the principal target of sedatives such as benzodiazepines and
15   barbiturates which act by enhancing GABA-mediated effects (Katzung, B.G. (1995) Basic and Clinical Pharmacology, 6th edition, Appleton & Lange, Norwalk, CT, pp. 338-339).

Two major classes of neurotransmitter transporters are essential to the function of the nervous system. The first class is uptake carriers in the plasma membrane of neurons and glial cells, which pump neurotransmitters from the extracellular space into the cell. This process relies on the $Na^+$
20   gradient across the plasma membrane, particularly the co-transport of $Na^+$. Two families of proteins have been identified. One family includes the transporters for GABA, monoamines such as noradrenaline, dopamine, serotonin, and amino acids such as glycine, and proline. Common structural components include twelve putative transmembrane α-helical domains, cytoplasmic N- and C- termini, and a large glycosylated extracellular loop separating transmembrane domains three and
25   four. This family of homologous proteins derives their energy from the co-transport of $Na^+$ and $Cl^-$ ions with the neurotransmitter into the cell ($Na^+/Cl^-$ neurotransmitter transporters). The second family includes transporters for excitatory amino acids such as glutamate. Common structural components include putative 6-10 transmembrane domains, cytoplasmic N- and C- termini, and glycosylations in the extracellular loops. The excitatory amino acid transporters are not dependent on $Cl^-$, and may
30   require intracellular $K^+$ ions ($Na^+/K^+$- neurotransmitter transporters) (Liu, Y. et al. (1999) Trends Cell Biol. 9:356-363).

The second class of neurotransmitter transporters is present in the vesicle membrane, and concentrates neurotransmitters from the cytoplasm into the vesicle, before exocytosis of the vesicular contents during synaptic transmission. Vesicular transport uses the electrochemical gradient across

the vesicular membrane generated by a H⁺-ATPase. Two families of proteins are involved in the transport of neurotransmitters into vesicles. One family uses primarily proton exchange to drive transport into secretory vesicles and includes the transporters for monoamines and acetylcholine. For example, the monoamine transporters exchange two luminal protons for each molecule of cytoplasmic

5    transmitter. The second family includes the GABA transporters, which relies on the positive charge inside synaptic vesicles. The two classes of vesicular transporters show no sequence similarity to each other and have structures distinct from those of the plasma membrane carriers (Schloss, P. et al. (1994) Curr. Opin. Cell Biol. 6:595-599; Liu, Y. et al. (1999) Trends Cell Biol. 9:356-363).

GABA is the predominant inhibitory neurotransmitter and is widely distributed in the

10   mammalian nervous system. GABA is cleared from the synaptic cleft by specific, high-affinity, Na⁺- and Cl⁻- dependent transporters, which are thought to be localized to both pre- and postsynaptic neurons, as well as to surrounding glial cells. At least four GABA transporters (GAT1–GAT4) have been cloned (Liu, Q.-R. et al. (1993) J. Biol. Chem. 268:2106–2112). Studies of [³H]-GABA uptake into cultured cells and plasma-membrane vesicles isolated from various tissues revealed considerable

15   differences in GABA transporter heterogeneity. GABA transporters exhibit differences in substrate affinity and specificity, distinct blocker pharmacologies, and different tissue localization. For example, the $K_m$ values of GABA uptake of the expressed GAT1 to GAT4 are 6, 79, 18, and 0.8 μM, respectively. In addition to transporting GABA, GAT2 also transports betaine; GAT3 and GAT4 also transport β-alanine and taurine. Pharmacological studies revealed that GABA transport by GAT1 and

20   GAT4 is more sensitive to 2,4-diaminobutyric acid and guavicine than that by GAT2 and GAT3. In situ hybridization showed that GAT1 and GAT4 expression is brain specific. GAT2 and GAT3 mRNAs were detected in tissues such as liver and kidney (Schloss, P. et al. (1994) Curr. Opin. Cell Biol. 6:595-599; Borden, L.A. (1996) Neurochem. Int. 29:335-356; Nelson, N. (1998) J. Neurochem. 71:1785-1803).

25   Human studies indicated that GABA transporter function is reduced in epileptic hippocampi. Decreased GABAergic neurotransmission has also been implicated in the pathophysiology of schizophrenia (Simpson, M.D. et al. (1992) Psychiatry Res. 42:273-282).

Diazepam binding inhibitor (DBI), also known as endozepine and acyl-Coenzyme (CoA)-binding protein, is an endogenous GABA receptor ligand which is thought to down-regulate the effects

30   of GABA. DBI binds medium- and long-chain acyl-CoA esters with very high affinity and may function as an intracellular carrier of acyl-CoA esters (*125950 Diazepam Binding Inhibitor; DBI, Online Mendelian Inheritance in Man (OMIM); PROSITE PDOC00686 Acyl-CoA-binding protein signature).

Glycine serves as one of the major inhibitory neurotransmitters in the mammalian nervous

system by activating chloride-channel receptors, which are members of a ligand-gated ion-channel superfamily (Betz, H. (1990) Neuron 5:383-392). Glycine also facilitates excitatory transmission through an allosteric activation of the N-methyl- D- aspartate (NMDA) receptor (Johnson, J.W. and P. Ascher (1987) Nature 325:529-531). Two forms of glycine transporters have been described to date,

5   GLYT 1 and GLYT 2. The two variants of GLYT1 (GLYT1 a/b) are generated by alternative splicing (Liu, Q.-R. et al. (1993) J. Biol. Chem. 268:22802-22808) and differ in their amino-terminal sequences. The proteins have similar affinity for glycine and both are sensitive to N-methyl-aminoacetic acid. *In situ* hybridization revealed that GLYT1a transcripts were detected in a both neural and non-neural tissues, whereas GLYT1b was detected only in neural tissues (Borowsky,

10   B. et al. (1993) Neuron 10:851-863). High levels of GLYT1a/b mRNA were found in hippocampus and cortex, suggesting a co-localization of this transporter with NMDA receptors and implying its involvement in the regulation of excitatory synaptic transmission. It is not clear whether GLYT1a is expressed in neurons, in glia or in both. In contrast, GLYT1b is found almost exclusively in fiber tracts, suggesting its localization in glial cells (Schloss, P. et al. (1994) Curr. Opin. Cell Biol.

15   6:595-599).

The second identified glycine transporter (GLYT2) differs from GLYT1a/b by its extended intracellular amino terminus. The predominant localization of its mRNA in brainstem and spinal cord and its insensitivity to N-methyl-aminoacetic acid suggests that GLYT2 terminates signal transduction at the strychnine-sensitive inhibitory glycine receptor. It has been proposed that, upon

20   depolarization of cells harboring GLYT1b, the transporter runs backwards and releases glycine to act as a neuromodulatory amino acid at the NMDA receptor (Attwell, D. and M. Bouvier (1992) Curr. Biol. 2:541-543). Such a $Ca^{2+}$-independent, non-vesicular release of neurotransmitters by reverse transport was demonstrated for glutamate and serotonin. This evidence suggests that the transmitter transporters may be important for both the initiation and termination of neurotransmitter action

25   (Schloss, P. et al. (1994) Curr. Opin. Cell Biol. 6:595-599).

The plasma membrane dopamine transporter (DAT) is essential for the reuptake of released dopamine from the synapse. Uptake of dopamine is temperature- and time-dependent, and is inhibited by a variety of compounds, such as cocaine. DAT- knockout mice have been shown to exhibit extreme hyperactivity and resistance to both cocaine and amphetamine, consistent with the

30   primary action of cocaine on DAT (Giros, B. et al. (1996) Nature 379:606-612). The perturbation of the tightly regulated DAT also predisposes neurons to damage by a variety of insults. Most notable is the selective degeneration of DAT-expressing dopamine nerve terminals in the striatum thought to underlie Parkinson's disease. DAT expression can predict the selective vulnerability of neuronal populations, which suggests that therapeutic strategies aimed at altering DAT function could have

35   significant benefits in a variety of disorders (Gary, W.M. et al. (1999) Trends Pharmacol. Sci.

20:424-429).

Creatine transporters are strongly related to transporters for GABA. The primary sequence identity between creatine transporter species homologs is very high (98-99%). Pharmacological characterization demonstrated high affinity creatine uptake (27-43 μM), which was blocked by

5   creatine analogs with high affinity. Creatine transporters are widely expressed in a variety of mammalian tissues, including brain, adrenal gland, intestine, colon, prostate, thymus, ovary, spleen, pancreas, placenta, umbilical cord, thyroid, tongue, pharnyx, vertebral discs, jaw, and nasal epithelium. Genetic mapping in the mouse localizes the creatine transporter to a region on the X chromosome in linkage conservation with the human region Xq28, the location of the genes for

10  several neuromuscular diseases (Nash, S.R. et al. (1994) Receptors Channels 2:165-174).

The substrates of a number of cDNA clones encoding proteins of the $Na^+$/$Cl^-$-dependent transporter families are still not identified. These are orphan transporters. Identification of the substrates for orphan transporters has been difficult because in situ hybridization and immunohistochemistry indicate that the transporters are synthesized by phenotypically different

15  neuronal populations, for example glutaminergic, GABAergic, histaminergic, or serotoninergic neurons. One of the transporters, NTT4 exhibits the highest homology to the creatine transporter. It differs structurally from other members of this family in having an unusually long loop between transmembranes seven and eight (Liu, Q.-R. et al. (1993) FEBS Lett. 315:114-118; Schloss, P. et al. (1994) Curr. Opin. Cell Biol. 6:595-599).

20  Glutamate is a major excitatory neurotransmitter in the mammalian central nervous system. Electrogenic ($Na^+$ / $K^+$)-coupled glutamate transporters, located in the plasma membranes of nerve terminals and glial cells, mediate removal of glutamate released at excitatory synapses and maintain extracellular concentrations below neurototoxic levels. Glutamate transporters achieve this process by co-transport with three sodium ions and one proton, followed by translocation of a potassium ion

25  in the opposite direction (Zerangue, N. and M.P. Kavanaugh (1996) Nature 383:634-637).

Glutamate transporters belong to a large family of transport proteins. The membrane topology of the glutamate transporter reveals six membrane-spanning helices in the N-terminal part of the proteins (Slotboom, D.J. et al. (1999) Microbiol. Mol. Biol. Rev. 63:293-307). The C-terminal half of the glutamate transporter is well conserved and constitutes a major part of the translocation

30  pathway and contains the binding sites for the substrate and co-transported ions (Zhang, Y. and B.I. Kanner (1999) Proc. Natl. Acad. Sci. USA 96:1710-1715).

Impaired re-uptake of synaptic glutamate, and a reduced expression of glutamate transporters have been found in the motor cortex of patients with amyotrophic lateral sclerosis (ALS). Inhibition of the synthesis of each glutamate transporter subtype using chronic antisense oligonucleotide

35  administration, *in vitro* and *in vivo*, selectively and specifically reduced the protein expression and

function of glutamate transporters. The loss of glial glutamate transporters produced elevated extracellular glutamate levels, neurodegeneration characteristic of excitotoxicity, and a progressive paralysis. The loss of the neuronal glutamate transporter did not elevate extracellular glutamate in the striatum but produced mild neurotoxicity and resulted in epilepsy (Rothstein, J.D. et al. (1996)

5     Neuron 16:675-686).

       The vesicular monoamine transporters (VMAT) package cytoplasmic monoamine neurotransmitters into secretory vesicles for regulated exocytotic release. VMAT acts as an electrogenic exchanger of protons and monoamines, using a proton electrochemical gradient. Two VMAT transporters, VMAT1 and VMAT2 have been cloned from rat, bovine and human. The

10     VMAT proteins possess twelve transmembrane segments, with both extremities lying on the cytoplasmic side. Examination of the subcellular localization of VMAT proteins indicates that they are associated with distinct vesicle populations in neurons and neuroendocrine cells. In rat, VMAT1 is expressed in the adrenal gland whereas VMAT2 is expressed in the brain. In contrast, the bovine adrenal gland expresses both VMAT1 and VMAT2 (Henry, J.-P. et al. (1994) J. Exp. Biol.

15     196:251-262.).

       Vesicular transport is inhibited by the antihypertensive drug reserpine and the related but more centrally acting drug tetrabenazine. The mechanism of transport and the biochemistry of VMAT have been analyzed with these drugs, using mainly the chromaffin granules from bovine adrenal glands as a source of transporters (Peter, D. et al. (1994) J. Biol. Chem. 269:7231-7237).

20     Human studies indicated that reserpine can cause a syndrome resembling depression, indicating the importance of vesicular transport activity for the control of mood and behavior. The psychostimulant amphetamine also disrupts the storage of amines in secretory vesicles, further indicating that alterations in vesicular monoamine transport can affect behavior (Sulzer, D. and S. Rayport (1990) Neuron 5:797-808).

25     Human diseases caused by defects in neurotransmitter transporters include schizophrenia, Tourette's syndrome, Parkinson's disease, brain ischemia, amyotrophic lateral scerlosis, depression, and epilepsy. For example, decreased GABAergic neurotransmission has been implicated in the pathophysiology of CNS disorders such as epilepsy and schizophrenia. Impaired re-uptake of synaptic glutamate, and a reduced expression of the glutamate transporter have been found in the

30     motor cortex of patients with amyotrophic lateral sclerosis (ALS). The loss of glial glutamate transporters produces elevated extracellular glutamate levels, neurodegeneration characteristic of excitotoxicity, and a progressive paralysis. The loss of neuronal glutamate transporters produces mild neurotoxicity and result in epilepsy (Rothstein, J.D. et al. (1996) Neuron 16:675-686).

       Transporters for dopamine, norepinephrine, and serotonin have particular significance as

35     targets for clinically relevant psychoactive agents including cocaine, antidepressants, and

amphetamines. Cocaine and antidepressants are transporter antagonists that act with varying degrees of specificity to enhance synaptic concentrations of amines by limiting clearance. Amphetamines enhance transporter mediated efflux in concert with a depletion of vesicular amine stores (Barker, E.L. and R.D. Blakely (1995) Psychopharmacology 28:321-333; Sulzer, D. and S. Rayport (1990)

5      Neuron 5:797-808; Wall, S.C. et al. (1995) Mol. Pharmacol. 47:544-550).

Each of over a trillion neurons in adult humans connects with over a thousand target cells (Tessier-Lavigne, M. et al. (1996) Science 274:1123-1133). These neuronal connections form during embryonic development. Each differentiating neuron sends out an axon tipped at the leading edge by a growth cone. Aided by molecular guidance cues, the growth cone migrates through the embryonic

10    environment to its synaptic target. Semaphorins are growth cone guidance signals that may function during embryogenesis by providing local signals to specify territories inaccessible to growing axons (Puschel, A.W. et al. (1995) Neuron 14:941-948).

The guidance of axons during development involves both positive and negative effects (*i.e.*, chemoattraction and chemorepulsion). The Slit family of proteins have been implicated in promoting

15    axon branching, elongation, and repulsion. Members of the Slit family have been identified in a variety of organisms, including insects, amphibians, birds, rodents and humans (Guthrie, S. (1999) Current Biology 9:R432-R435). Slit proteins appear to be ligands for the repulsive guidance receptor, Roundabout (Robo); however, Slit also cause elongation in some assays. A post-translationally processed form of Slit appears to the active form of the protein (Guthrie, S. supra and Brose, K. et al.

20    (1999) Cell 96:795-806).

Axon growth is also guided in part by contact-mediated mechanisms involving cell surface and extracellular matrix (ECM) molecules. Many ECM molecules, including fibronectin, vitronectin, members of the laminin, tenascin, collagen, and thrombospondin families, and a variety of proteoglycans, can act either as promoters or inhibitors of neurite outgrowth and extension (Tessier-

25    Lavigne et al., supra). Receptors for ECM molecules include integrins, immunoglobulin superfamily members, and proteoglycans. ECM molecules and their receptors have also been implicated in the adhesion, maintenance, and differentiation of neurons (Reichardt, L.F. et al. (1991) Ann. Rev. Neurosci. 14:531-571). The proteoglycan testican is localized to the post-synaptic area of pyramidal cells of the hippocampus and may play roles in receptor activity, neuromodulation, synaptic plasticity,

30    and neurotransmission (Bonnet, F. et al. (1996) J. Biol. Chem. 271:4373-4380).

Other nervous system-associated proteins have roles in neuron signaling, cell adhesion, nerve regeneration, axon guidance, and neurogenesis. The neurexophilins are neuropeptide-like proteins which are proteolytically processed after synthesis. They are ligands for the neuron-specific cell surface proteins, the α-neurexins. Neurexophilins and neurexins may participate in a neuron signaling

pathway (Missler, M. and T.C. Sudhof (1998) J. Neurosci. 18:3630-3638; Missler, M. et al. (1998) J. Biol. Chem. 273:34716-34723). Ninjurin is a neuron cell surface protein which plays a role in cell adhesion and in nerve regeneration following injury. Ninjurin is up-regulated after nerve injury in dorsal root ganglion neurons and in Schwann cells (*602062 Ninjurin; NINJ1 OMIM; Araki, T. and

5      Milbrandt, J. (1996) Neuron 17:353-361). Mammalian Numb is a phosphotyrosine-binding (PTB) domain-containing protein which may be involved in cortical neurogenesis and cell fate decisions in the mammalian nervous system. Numb's binding partner, the LNX protein, contains four PDZ domains and a ring finger domain and may participate in a signaling pathway involving Numb. PDZ domains have been found in proteins which act as adaptors in the assembly of multifunctional protein

10     complexes involved in signaling events at surfaces of cell membranes (Ponting, C.P. (1997) Bioessays 19:469-479). LNX contains a tyrosine phosphorylation site which may be important for the binding of other PTB-containing proteins such as SHC, an adaptor protein which associates with tyrosine-phosphorylated growth factor receptors and downstream effectors (Dho, S.E. et al. (1998) J. Biol. Chem. 273:9179-9187).

15          Another family of molecules that appear to be important for neurotransmission is the choline-transporter-like CTL1 proteins. The prototypic CTL1 was identified in yeast as a suppressor of a choline transport mutation; however, mammalian homologues have been identified. The proteins comprise approximately ten putative transmembrane domains in addition to transporter-like motifs but do not appear to be canonical choline transporters. Choline transport is important to neurotransmission

20     because choline is a precursor of acetylcholine, required in abundance by cholinergic neurons (O'Regan, S. et al. (2000) Proc. Natl. Acad. Sci. U.S.A. 97:1835-40).

            The discovery of new neurotransmitter transporters and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of transport, neurological and psychiatric disorders, and in the assessment of the effects

25     of exogenous compounds on the expression of nucleic acid and amino acid sequences of neurotransmitter transporters.


## SUMMARY OF THE INVENTION

            The invention features purified polypeptides, neurotransmitter transporters, referred to

30     collectively as "NTT" and individually as "NTT-1," "NTT-2," "NTT-3," "NTT-4," "NTT-5," and "NTT-6,". In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a

biologically active fragment of a polypeptide having an amino acid sequence selected from the group

consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid

sequence selected from the group consisting of SEQ ID NO:1-6. In one alternative, the invention

provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-6.

5          The invention further provides an isolated polynucleotide encoding a polypeptide selected from

the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group

consisting of SEQ ID NO:1-6, b) a naturally occurring polypeptide comprising an amino acid sequence

at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the

10       group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an

amino acid sequence selected from the group consisting of SEQ ID NO:1-6. In one alternative, the

polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-6. In another

alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:7-12.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter

15       sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting

of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID

NO:1-6, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical

to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically

active fragment of a polypeptide having an amino acid sequence selected from the group consisting of

20       SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence

selected from the group consisting of SEQ ID NO:1-6. In one alternative, the invention provides a cell

transformed with the recombinant polynucleotide. In another alternative, the invention provides a

transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group

25       consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of

SEQ ID NO:1-6, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90%

identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a

biologically active fragment of a polypeptide having an amino acid sequence selected from the group

consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid

30       sequence selected from the group consisting of SEQ ID NO:1-6. The method comprises a) culturing a

cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a

recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide

encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a

polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid

5      sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90%

10     identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample,

15     said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the

20     polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of

25     said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID

30     NO:7-12, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b)

14

detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from

5    the group consisting of SEQ ID NO:1-6, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and a pharmaceutically

10   acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional NTT, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an

15   agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and

20   d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of

25   treating a disease or condition associated with decreased expression of functional NTT, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a naturally occurring

30   polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. The method comprises a) exposing a sample comprising the

polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional NTT, comprising
5    administering to a patient in need of such treatment the composition.

        The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a naturally occurring polypeptide cmoprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the
10   group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a
15   compound that specifically binds to the polypeptide.

        The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence
20   selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the
25   activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

        The invention further provides a method for screening a compound for effectiveness in
30   altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:7-12, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

        The invention further provides a method for assessing toxicity of a test compound, said

method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, ii) a

5      naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological

10     sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, ii) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide

15     of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated

20     biological sample is indicative of toxicity of the test compound.


## BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

25     Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of

30     the polypeptides.

Table 4 lists the cDNA and genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

5

## DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing

10     particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a

15     reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to

20     practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

25     **DEFINITIONS**

"NTT" refers to the amino acid sequences of substantially purified NTT obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of

30     NTT. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of NTT either by directly interacting with NTT or by acting on components of the biological pathway in which NTT participates.

An "allelic variant" is an alternative form of the gene encoding NTT. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in

18

polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in

5     a given sequence.

"Altered" nucleic acid sequences encoding NTT include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as NTT or a polypeptide with at least one functional characteristic of NTT. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of

10    · the polynucleotide encoding NTT, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding NTT. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent NTT. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility,

15    hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of NTT is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains

20    having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring

25    protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

30    The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of NTT. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of NTT either by directly interacting with NTT or by acting on components of the biological pathway in which NTT participates.

19

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind NTT polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used

5      to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that

10     makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

15     The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having

20     modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the

25     designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic NTT, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

30     "Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or

amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding NTT or fragments of NTT may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be

5    deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from

10   one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least

15   interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

|        | Original Residue | Conservative Substitution |
|--------|------------------|---------------------------|
| 20     | Ala              | Gly, Ser                  |
|        | Arg              | His, Lys                  |
|        | Asn              | Asp, Gln, His             |
|        | Asp              | Asn, Glu                  |
|        | Cys              | Ala, Ser                  |
| 25     | Gln              | Asn, Glu, His             |
|        | Glu              | Asp, Gln, His             |
|        | Gly              | Ala                       |
|        | His              | Asn, Arg, Gln, Glu        |
|        | Ile              | Leu, Val                  |
| 30     | Leu              | Ile, Val                  |
|        | Lys              | Arg, Gln, Glu             |
|        | Met              | Leu, Ile                  |
|        | Phe              | His, Met, Leu, Trp, Tyr   |
|        | Ser              | Cys, Thr                  |
| 35     | Thr              | Ser, Val                  |
|        | Trp              | Phe, Tyr                  |
|        | Tyr              | His, Phe, Trp             |
|        | Val              | Ile, Leu, Thr             |

40       Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation,

(b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

5      The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological

10     function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such

15     comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

A "fragment" is a unique portion of NTT or the polynucleotide encoding NTT which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a

20     fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected

25     from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:7-12 comprises a region of unique polynucleotide sequence that

30     specifically identifies SEQ ID NO:7-12, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:7-12 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:7-12 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:7-12 and the region of SEQ ID NO:7-12 to which the fragment corresponds are routinely

determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-6 is encoded by a fragment of SEQ ID NO:7-12. A fragment of SEQ ID NO:1-6 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-6. For example, a fragment of SEQ ID NO:1-6 is useful as an immunogenic peptide for the

5    development of antibodies that specifically recognize SEQ ID NO:1-6. The precise length of a fragment of SEQ ID NO:1-6 and the region of SEQ ID NO:1-6 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon

10   (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to

15   the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default

20   parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows:

25   Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search

30   Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2

Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2

Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The

"BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST

programs are commonly used with gap and other parameters set to default settings. For example, to

5      compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version

2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

       *Matrix: BLOSUM62*

       *Reward for match: 1*

       *Penalty for mismatch: -2*

10            *Open Gap: 5 and Extension Gap: 2 penalties*

       *Gap x drop-off: 50*

       *Expect: 10*

       *Word Size: 11*

       *Filter: on*

15     Percent identity may be measured over the length of an entire defined sequence, for example, as

defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over

the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at

least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such

lengths are exemplary only, and it is understood that any fragment length supported by the sequences

20     shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which

percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode

similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in

a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences

25     that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to

the percentage of residue matches between at least two polypeptide sequences aligned using a

standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment

methods take into account conservative amino acid substitutions. Such conservative substitutions,

30     explained in more detail above, generally preserve the charge and hydrophobicity at the site of

substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters

of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment

program (described and referenced above). For pairwise alignments of polypeptide sequences using

24

CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

5          Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

          *Matrix: BLOSUM62*

          *Open Gap: 11 and Extension Gap: 1 penalties*

10        *Gap x drop-off: 50*

          *Expect: 10*

          *Word Size: 3*

          *Filter: on*

          Percent identity may be measured over the length of an entire defined polypeptide sequence, for

15   example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to

20   describe a length over which percentage identity may be measured.

          "Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

          The term "humanized antibody" refers to an antibody molecule in which the amino acid

25   sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

          "Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity.

30   Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and

may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

5          Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ($T_m$) for the specific sequence at a defined ionic strength and pH. The $T_m$ is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating $T_m$ and conditions

10  for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour.

15  Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances,

20  such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid

25  sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., $C_0t$ or $R_0t$ analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

30          The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular

and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of NTT which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of
5    NTT which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.
10    The term "modulate" refers to a change in the activity of NTT. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of NTT.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or
15    synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding
20    sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs
25    preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an NTT may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by
30    cell type depending on the enzymatic milieu of NTT.

"Probe" refers to nucleic acid sequences encoding NTT, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are

short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

5      Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the

10     specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR

15     Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such

20     purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South

25     West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for

30     microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences.

Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary

5    polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the

10   artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

15   Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs).

20   Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and

25   other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

30   The term "sample" is used in its broadest sense. A sample suspected of containing NTT, nucleic acids encoding NTT, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a

29

protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope

5    A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which

10   they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers,

15   microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient

20   cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells"

25   includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic

30   acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The

transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention

5   into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999)

10  set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant

15  identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic

20  variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at

25  least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a

30  certain defined length of one of the polypeptides.

## THE INVENTION

The invention is based on the discovery of new human neurotransmitter transporters (NTT), the polynucleotides encoding NTT, and the use of these compositions for the diagnosis, treatment, or

31

prevention of transport, neurological and psychiatric disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted

5   by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by
10  BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank

15  homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3

20  shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable

25  databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are neurotransmitter transporters. For example, SEQ ID NO:4 is 95% identical to human taurine transporter (GenBank ID g799339) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0,

30  which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:4 also contains a sodium:neurotransmitter symporter family (SNF) signature as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:4 is an SNF

neurotansmitter transporter. In an alternate example, SEQ ID NO:5 is 94% identical to human
dopamine transporter (GenBank ID g181656) as determined by the Basic Local Alignment Search Tool
(BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of
obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:5 also contains a

5      sodium neurotransporter symporter family domain as determined by searching for statistically
significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein
family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide
further corroborative evidence that SEQ ID NO:5 is a dopamine transporter. In an alternate example,
SEQ ID NO:6 is 90% identical to rat neurotransmitter transporter rB21a (GenBank ID g914028;

10     Smith, K.E. et al. (1995) FEBS Lett. 357:86-92) as determined by the Basic Local Alignment Search
Tool (BLAST). (See Table 2.) The BLAST probability score is 5.7e-300, which indicates the
probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also
contains sodium:neurotransmitter symporter family domains as determined by searching for statistically
significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein

15     family domains. (See Table 3.) Data from BLIMPS, PROFILESCAN, and BLAST analyses provide
further corroborative evidence that SEQ ID NO:6 is a sodium:neurotransmitter symporter. SEQ ID
NO:1, SEQ ID NO:2, and SEQ ID NO:3 were analyzed and annotated in a similar manner. The
algorithms and parameters for the analysis of SEQ ID NO:1-6 are described in Table 7.

         As shown in Table 4, the full length polynucleotide sequences of the present invention were
20     assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any
combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence
identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide
consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention.
Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of

25     the polynucleotide sequences which are useful, for example, in hybridization or amplification
technologies that identify SEQ ID NO:7-12 or that distinguish between SEQ ID NO:7-12 and related
polynucleotide-sequences. Column 5 shows identification numbers corresponding to cDNA sequences,
coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of
both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide

30     sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3')
positions of the cDNA and genomic sequences in column 5 relative to their respective full length
sequences.

         The identification numbers in Column 5 of Table 4 may refer specifically, for example, to
Incyte cDNAs along with their corresponding cDNA libraries. For example,

5407696F6 is the identification number of an Incyte cDNA sequence, and BRAMNOT01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 70501086V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g1748013) which contributed to the assembly of

5    the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. (See Example V.) Alternatively, the identification numbers

10   in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. For example, FL7472800_g7108462_g181656 is the identification number of a "stretched" sequence, with 7472800 being the Incyte project identification number, g7108462 being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, and g181656 being the GenBank identification number of the

15   nearest GenBank protein homolog. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte

20   cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses NTT variants. A preferred NTT variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence

25   identity to the NTT amino acid sequence, and which contains at least one functional or structural characteristic of NTT.

The invention also encompasses polynucleotides which encode NTT. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:7-12, which encodes NTT. The polynucleotide sequences of SEQ

30   ID NO:7-12, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding NTT. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least

34

about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence
encoding NTT. A particular aspect of the invention encompasses a variant of a polynucleotide
sequence comprising a sequence selected from the group consisting of SEQ ID NO:7-12 which has at
least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide

5    sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:7-12.
Any one of the polynucleotide variants described above can encode an amino acid sequence which
contains at least one functional or structural characteristic of NTT.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic
code, a multitude of polynucleotide sequences encoding NTT, some bearing minimal similarity to the

10   polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the
invention contemplates each and every possible variation of polynucleotide sequence that could be made
by selecting combinations based on possible codon choices. These combinations are made in
accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally
occurring NTT, and all such variations are to be considered as being specifically disclosed.

15   Although nucleotide sequences which encode NTT and its variants are generally capable of
hybridizing to the nucleotide sequence of the naturally occurring NTT under appropriately selected
conditions of stringency, it may be advantageous to produce nucleotide sequences encoding NTT or its
derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring
codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a

20   particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons
are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding NTT
and its derivatives without altering the encoded amino acid sequences include the production of RNA
transcripts having more desirable properties, such as a greater half-life, than transcripts produced from
the naturally occurring sequence.

25   The invention also encompasses production of DNA sequences which encode NTT and NTT
derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic
sequence may be inserted into any of the many available expression vectors and cell systems using
reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into
a sequence encoding NTT or any fragment thereof.

30   Also encompassed by the invention are polynucleotide sequences that are capable of
hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID
NO:7-12 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and
S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol.

152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of

5    DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV),

10   PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols

15   in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding NTT may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed,

20   restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids

25   Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences

30   are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences,

Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been
5     size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the
10    size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire
15    process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode NTT may be cloned in recombinant DNA molecules that direct expression of NTT, or
20    fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express NTT.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter NTT-encoding sequences for a variety of purposes including, but not
25    limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

30    The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of NTT, such as its biological or enzymatic activity or its ability to

bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of

5    DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby

10   maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding NTT may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively,

15   NTT itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence

20   of NTT, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.)

25   The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active NTT, the nucleotide sequences encoding NTT or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a

30   suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding NTT. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding NTT. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where

sequences encoding NTT and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the

5    vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

       Methods which are well known to those skilled in the art may be used to construct expression

10   vectors containing sequences encoding NTT and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

15     A variety of expression vector/host systems may be utilized to contain and express sequences encoding NTT. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or

20   tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New

25   York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci.

30   USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

       In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding NTT. For example, routine cloning,

subcloning, and propagation of polynucleotide sequences encoding NTT can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding NTT into the vector's multiple cloning site disrupts the lacZ gene, allowing a colorimetric screening procedure for identification of transformed

5 bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of NTT are needed, e.g. for the production of antibodies, vectors which direct high level expression of NTT may be used. For example, vectors containing the

10 strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of NTT. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into

15 the host genome for stable propagation. (See, e.g., Ansubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of NTT. Transcription of sequences encoding NTT may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311).

20 Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp.

25 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding NTT may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain

30 infective virus which expresses NTT in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

5        For long term production of recombinant proteins in mammalian systems, stable expression of NTT in cell lines is preferred. For example, sequences encoding NTT can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being

10      switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include,

15      but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to

20      chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß

25      glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is

30      also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding NTT is inserted within a marker gene sequence, transformed cells containing sequences encoding NTT can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding NTT under the control of a single

promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding NTT and that express NTT may be identified by a variety of procedures known to those of skill in the art. These procedures

5 include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of NTT using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include

10 enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on NTT is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et

15 al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding NTT include oligolabeling,

20 nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding NTT, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available

25 kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding NTT may be cultured under

30 conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode NTT may be designed to contain signal sequences which direct secretion of NTT through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the

5    protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

10   In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding NTT may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric NTT protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of NTT activity. Heterologous protein and peptide moieties

15   may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, c-myc, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins,

20   respectively. FLAG, c-myc, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the NTT encoding sequence and the heterologous protein sequence, so that NTT may be cleaved away from the heterologous moiety following purification. Methods for fusion protein

25   expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled NTT may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or

30   SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

NTT of the present invention or fragments thereof may be used to screen for compounds that specifically bind to NTT. At least one and up to a plurality of test compounds may be screened for

specific binding to NTT. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of NTT, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a

5    natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which NTT binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express NTT, either as a secreted

10   protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing NTT or cell membrane fractions which contain NTT are then contacted with a test compound and binding, stimulation, or inhibition of activity of either NTT or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is

15   detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with NTT, either in solution or affixed to a solid support, and detecting the binding of NTT to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical

20   libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

NTT of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of NTT. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for NTT

25   activity, wherein NTT is combined with at least one test compound, and the activity of NTT in the presence of a test compound is compared with the activity of NTT in the absence of the test compound. A change in the activity of NTT in the presence of the test compound is indicative of a compound that modulates the activity of NTT. Alternatively, a test compound is combined with an in vitro or cell-free system comprising NTT under conditions suitable for NTT activity, and the assay is performed. In

30   either of these assays, a test compound which modulates the activity of NTT may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding NTT or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES)

cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest

5       disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330).

10      Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

        Polynucleotides encoding NTT may also be manipulated in vitro in ES cells derived from

15      human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

        Polynucleotides encoding NTT can also be used to create "knockin" humanized animals (pigs)

20      or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding NTT is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal

25      inbred to overexpress NTT, e.g., by secreting NTT in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

**THERAPEUTICS**

        Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of NTT and neurotransmitter transporters. In addition, the expression of NTT is

30      closely associated with brain, neurological, secretory, and tumor tissue. Therefore, NTT appears to play a role in transport, neurological and psychiatric disorders. In the treatment of disorders associated with increased NTT expression or activity, it is desirable to decrease the expression or activity of NTT. In the treatment of disorders associated with decreased NTT expression or activity, it is desirable to increase the expression or activity of NTT.

Therefore, in one embodiment, NTT or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NTT. Examples of such disorders include, but are not limited to, a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy,

5  Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer; cardiac disorders associated with transport, e.g., angina, bradyarrythmia,

10  tachyarrythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis; neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia; and other

15  disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn

20  syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other

25  demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis,

30  encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental

35  disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD),

46

akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a psychiatric disorder such as acute stress disorder, alcohol dependence, amphetamine dependence, anorexia nervosa, antisocial personality disorder,

5     attention-deficit hyperactivity disorder, autistic disorder, anxiety, avoidant personality disorder, bipolar disorder, borderline personality disorder, brief psychotic disorder, bulimia nervosa, cannabis dependence, cocaine dependence, conduct disorder, cyclothymic disorder, delirium, delusional disorder, dementia, dependent personality disorder, depression, dysthymic disorder, hallucinogen dependence, histrionic personality disorder, inhalant dependence, manic depression, multi-infarct

10    dementia, narcissistic personality disorder, nicotine dependence, obsessive-compulsive disorder, opioid dependence, oppositional defiant disorder, panic disorder, paranoid personality disorder, phencyclidine dependence, phobia, posttraumatic stress disorder, schizoaffective disorder, schizoid personality disorder, schizophrenia, sedative dependence, separation anxiety disorder, and sleep disorder.

15          In another embodiment, a vector capable of expressing NTT or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NTT including, but not limited to, those described above.

            In a further embodiment, a composition comprising a substantially purified NTT in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder

20    associated with decreased expression or activity of NTT including, but not limited to, those provided above.

            In still another embodiment, an agonist which modulates the activity of NTT may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NTT including, but not limited to, those listed above.

25          In a further embodiment, an antagonist of NTT may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of NTT. Examples of such disorders include, but are not limited to, those transport, neurological and psychiatric disorders described above. In one aspect, an antibody which specifically binds NTT may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to

30    cells or tissues which express NTT.

            In an additional embodiment, a vector expressing the complement of the polynucleotide encoding NTT may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of NTT including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination

5    of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of NTT may be produced using methods which are generally known in the art. In particular, purified NTT may be used to produce antibodies or to screen libraries of pharmaceutical

10    agents to identify those which specifically bind NTT. Antibodies to NTT may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

15    For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with NTT or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols,

20    polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and _Corynebacterium parvum_ are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to NTT have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are

25    identical to a portion of the amino acid sequence of the natural protein. Short stretches of NTT amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to NTT may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited

30    to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda,

5     S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce NTT-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

10    Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for NTT may also be generated. For

15    example, such fragments include, but are not limited to, F(ab')$_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

20    Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between NTT and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two

25    non-interfering NTT epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for NTT. Affinity is expressed as an association constant, $K_a$, which is defined as the molar concentration of NTT-antibody complex divided by the

30    molar concentrations of free antigen and free antibody under equilibrium conditions. The $K_a$ determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple NTT epitopes, represents the average affinity, or avidity, of the antibodies for NTT. The $K_a$ determined for a preparation of monoclonal antibodies, which are monospecific for a particular NTT epitope, represents a true measure of affinity. High-affinity antibody preparations with $K_a$ ranging from about $10^9$ to $10^{12}$

49

L/mole are preferred for use in immunoassays in which the NTT-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with $K_a$ ranging from about $10^6$ to $10^7$ L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of NTT, preferably in active form, from the antibody (Catty, D. (1988) Antibodies,

5 Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg

10 specific antibody/ml, is generally employed in procedures requiring precipitation of NTT-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding NTT, or any fragment or

15 complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding NTT. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding NTT. (See,

20 e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g.,

25 Slater, J.E. et al. (1998) J. Allergy Cli. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other

30 systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding NTT may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency

(e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked

inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined

immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency

(Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475),

5      cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene

Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial

hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal,

R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii)

express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated

10     cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g.,

against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988)

Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399),

hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides

brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the

15     case where a genetic deficiency in NTT expression or regulation causes disease, the expression of NTT

from an appropriate population of transduced cells may alleviate the clinical manifestations caused by

the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in NTT

are treated by constructing mammalian expression vectors encoding NTT and introducing these vectors

20     by mechanical means into NTT-deficient cells. Mechanical transfer technologies for use with cells in

vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle

delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of

DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics,

Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

25     Expression vectors that may be effective for the expression of NTT include, but are not limited

to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA),

PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF,

PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). NTT may be expressed

using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus

30     (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the

tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA

89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998)

Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the

ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the

FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding NTT from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID
5   TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these
10  standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to NTT expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding NTT under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive
15  element (RRE) along with additional retrovirus cis-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for
20  receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a
25  method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4$^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al.
30  (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding NTT to cells which have one or more genetic abnormalities with respect to the expression of NTT. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile

52

for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr.

5   19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding NTT to target cells which have one or more genetic abnormalities with respect to the expression of NTT. The use of herpes simplex virus (HSV)-based vectors may be especially

10  valuable for introducing NTT to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413

15  to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV

20  vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

25  In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding NTT to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This

30  subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for NTT into the alphavirus genome in place of the capsid-coding region results in the production of a large number of NTT-coding RNAs and the synthesis of high levels of NTT in vector transduced cells. While alphavirus infection is typically

associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of NTT into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding NTT.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding NTT. Such DNA sequences may be incorporated into a wide variety of vectors with suitable

RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends

5    of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

10       An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding NTT. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular

15   chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased NTT expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding NTT may be therapeutically useful, and in the treament of disorders associated with

20   decreased NTT expression or activity, a compound which specifically promotes expression of the polynucleotide encoding NTT may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in

25   altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding NTT is exposed to at least one test compound thus obtained. The sample

30   may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding NTT are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding NTT. The amount of hybridization may be quantified, thus forming the

35   basis for a comparison of the expression of the polynucleotide both with and without exposure to one

or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins,

5    D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice,

10   T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved

15   using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

20   An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of NTT,

25   antibodies to NTT, and mimetics, agonists, antagonists, or inhibitors of NTT.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

30   Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled

the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

5      Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising NTT or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the

10     macromolecule. Alternatively, NTT or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell

15     culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example NTT or

20     fragments thereof, antibodies of NTT, and agonists, antagonists or inhibitors of NTT, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the $ED_{50}$ (the dose therapeutically effective in 50% of the population) or $LD_{50}$ (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the

25     therapeutic index, which can be expressed as the $LD_{50}/ED_{50}$ ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the $ED_{50}$ with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the

30     patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time

57

and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 $\mu$g to 100,000 $\mu$g, up to a total dose of
5    about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

10   **DIAGNOSTICS**

In another embodiment, antibodies which specifically bind NTT may be used for the diagnosis of disorders characterized by expression of NTT, or in assays to monitor patients being treated with NTT or agonists, antagonists, or inhibitors of NTT. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for NTT include
15   methods which utilize the antibody and a label to detect NTT in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring NTT, including ELISAs, RIAs, and FACS, are known in
20   the art and provide a basis for diagnosing altered or abnormal levels of NTT expression. Normal or standard values for NTT expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to NTT under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of NTT expressed in subject,
25   control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding NTT may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and
30   quantify gene expression in biopsied tissues in which expression of NTT may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of NTT, and to monitor regulation of NTT levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding NTT or closely related molecules may be used to

identify nucleic acid sequences which encode NTT. The specificity of the probe, whether it is made
from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a
conserved motif, and the stringency of the hybridization or amplification will determine whether the
probe identifies only naturally occurring sequences encoding NTT, allelic variants, or related sequences.

5        Probes may also be used for the detection of related sequences, and may have at least 50%
sequence identity to any of the NTT encoding sequences. The hybridization probes of the subject
invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:7-12 or from
genomic sequences including promoters, enhancers, and introns of the NTT gene.

         Means for producing specific hybridization probes for DNAs encoding NTT include the
10   cloning of polynucleotide sequences encoding NTT or NTT derivatives into vectors for the production
     of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to
     synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the
     appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups,
     for example, by radionuclides such as $^{32}P$ or $^{35}S$, or by enzymatic labels, such as alkaline phosphatase
15   coupled to the probe via avidin/biotin coupling systems, and the like.

         Polynucleotide sequences encoding NTT may be used for the diagnosis of disorders associated
     with expression of NTT. Examples of such disorders include, but are not limited to, a transport
     disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's
     muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus,
20   diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic
     periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia
     gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral
     neoplasms, prostate cancer; cardiac disorders associated with transport, e.g., angina, bradyarrythmia,
     tachyarrythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline
25   myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy,
     ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis;
     neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder,
     dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia; and other
     disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal
30   neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery
     stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter,
     Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome,
     hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn
     syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a

59

neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other

5   demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis,

10  encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental

15  disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a psychiatric disorder such as acute stress disorder, alcohol dependence, amphetamine dependence, anorexia nervosa, antisocial personality disorder,

20  attention-deficit hyperactivity disorder, autistic disorder, anxiety, avoidant personality disorder, bipolar disorder, borderline personality disorder, brief psychotic disorder, bulimia nervosa, cannabis dependence, cocaine dependence, conduct disorder, cyclothymic disorder, delirium, delusional disorder, dementia, dependent personality disorder, depression, dysthymic disorder, hallucinogen dependence, histrionic personality disorder, inhalant dependence, manic depression, multi-infarct

25  dementia, narcissistic personality disorder, nicotine dependence, obsessive-compulsive disorder, opioid dependence, oppositional defiant disorder, panic disorder, paranoid personality disorder, phencyclidine dependence, phobia, posttraumatic stress disorder, schizoaffective disorder, schizoid personality disorder, schizophrenia, sedative dependence, separation anxiety disorder, and sleep disorder. The polynucleotide sequences encoding NTT may be used in Southern or northern analysis,

30  dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered NTT expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding NTT may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide

35  sequences encoding NTT may be labeled by standard methods and added to a fluid or tissue sample

from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding NTT in the sample indicates

5    the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of NTT, a normal or standard profile for expression is established. This may be accomplished by combining body

10   fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding NTT, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from

15   patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from

20   successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual

25   clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding NTT may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or

30   produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding NTT, or a fragment of a polynucleotide complementary to the polynucleotide encoding NTT, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding NTT may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism

5   (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding NTT are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-

10  denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence

15  variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of NTT include radiolabeling or

20  biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid

25  quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify

30  genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective

62

treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, NTT, fragments of NTT, or antibodies specific for NTT may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein
5     interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a
10   given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present
15   invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.
20          Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity
25   (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression
30   provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not

necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed
5  gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present
10  invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global
15  pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is
20  achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is
25  generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry.
30  The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for NTT to quantify the levels of NTT expression. In one embodiment, the antibodies are used as elements on a microarray, and

protein expression levels are quantified by exposing the microarray to the sample and detecting the

levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111;

Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of

methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-

5     reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should

be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation

between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J.

Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the

10    analysis of compounds which do not significantly affect the transcript image, but which alter the

proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid

degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological

sample containing proteins with the test compound. Proteins that are expressed in the treated biological

15    sample are separated so that the amount of each protein can be quantified. The amount of each protein

is compared to the amount of the corresponding protein in an untreated biological sample. A difference

in the amount of protein between the two samples is indicative of a toxic response to the test compound

in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the

individual proteins and comparing these partial sequences to the polypeptides of the present invention.

20    In another embodiment, the toxicity of a test compound is assessed by treating a biological

sample containing proteins with the test compound. Proteins from the biological sample are incubated

with antibodies specific to the polypeptides of the present invention. The amount of protein recognized

by the antibodies is quantified. The amount of protein in the treated biological sample is compared with

the amount in an untreated biological sample. A difference in the amount of protein between the two

25    samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g.,

Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci.

USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al.

(1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-

30    2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well

known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999)

Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding NTT may be used to

generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either

coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a

5    chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic

10   linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map

15   data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding NTT on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as

20   linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized

25   by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

30   In another embodiment of the invention, NTT, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between NTT and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with NTT, or fragments thereof, and

5    washed. Bound NTT is then detected by methods well known in the art. Purified NTT can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding NTT specifically compete with a test compound for binding NTT. In this

10   manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with NTT.

In additional embodiments, the nucleotide sequences which encode NTT may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such

15   properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

20   The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/205,518, U.S. Ser. No. 60/213,956, U.S. Ser. No. 60/215,105, U.S. Ser. No. 60/218,947, and U.S. Ser. No. 60/221,488 are expressly incorporated by reference herein.


## EXAMPLES

25   I.    **Construction of cDNA Libraries**

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol

30   and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated

using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

5          In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic

10    oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g.,

15    PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

20    II.     **Isolation of cDNA Clones** ·

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL

25    8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal

30    cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III.    **Sequencing and Analysis**

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the

5    MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI

10    PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

15    The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS,

20    DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank

25    cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length

30    polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using

MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and
LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are
generated using default parameters specified by the CLUSTAL algorithm as incorporated into the
MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent
5    identity between aligned sequences.

         Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of
Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold
parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second
column provides brief descriptions thereof, the third column presents appropriate references, all of
10   which are incorporated by reference herein in their entirety, and the fourth column presents, where
applicable, the scores, probability values, and other parameters used to evaluate the strength of a match
between two sequences (the higher the score or the lower the probability value, the greater the identity
between two sequences).

         The programs described above for the assembly and analysis of full length polynucleotide and
15   polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID
NO:7-12. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and
amplification technologies are described in Table 4, column 4.

**IV.     Identification and Editing of Coding Sequences from Genomic DNA**

         Putative neurotransmitter transporters were initially identified by running the Genscan gene
20   identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a
general-purpose gene identification program which analyzes genomic DNA sequences from a variety of
organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin
(1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an
assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a
25   FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for
Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA
sequences encode neurotransmitter transporters, the encoded polypeptides were analyzed by querying
against PFAM models for neurotransmitter transporters. Potential neurotransmitter transporters were
also identified by homology to Incyte cDNA sequences that had been annotated as neurotransmitter
30   transporters. These selected Genscan-predicted sequences were then compared by BLAST analysis to
the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then
edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by
Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or
public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription.

When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were

5   derived entirely from edited or unedited Genscan-predicted coding sequences.

V.      Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

        Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped

10  to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in

15  the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the

20  longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit

25  from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

        Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public

30  databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the

chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to

5      determine whether it contained a complete gene.

VI.      **Chromosomal Mapping of NTT Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:7-12 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched

10     SEQ ID NO:7-12 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment

15     of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in

20     humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified

25     disease genes map within or in proximity to the intervals indicated above.

VII.      **Analysis of Polynucleotide Expression**

. Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995)

30     supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer

72

search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \text{ x minimum } \{\text{length(Seq. 1), length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding NTT are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding NTT. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII.   Extension of NTT Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using

5    OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension

10   was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing $Mg^{2+}$, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme

15   (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times;

20   Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II

25   (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and

30   sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs,

74

and transfected into competent <u>E. coli</u> cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham
5    Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2,
10   v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides
15   designed for such extension, and an appropriate genomic library.

**IX.    Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:7-12 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments.
20   Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 $\mu$Ci of [$\gamma$-$^{32}$P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing $10^7$ counts per
25   minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16
30   hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

**X.    Microarrays**

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested

5    substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470;

10   Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array

15   elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of

20   complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

### Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and

25   poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with

30   GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc.

(CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

5  **Microarray Preparation**

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5

10  μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR

15  Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average

20  concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate

25  buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

**Hybridization**

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample

30  mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm$^2$ coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC,

35  0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

77

## Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is
5 focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially.
10 Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source,
15 although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples
20 from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital
25 (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping
30 emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

### XI.     Complementary Polynucleotides

Sequences complementary to the NTT-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring NTT. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with

5     smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of NTT. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the NTT-encoding transcript.

10    ### XII.     Expression of NTT

Expression and purification of NTT is achieved using bacterial or virus-based expression systems. For expression of NTT in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the

15    T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express NTT upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of NTT in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant <u>Autographica californica</u> nuclear polyhedrosis virus (AcMNPV), commonly known as

20    baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding NTT by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect <u>Spodoptera frugiperda</u> (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional

25    genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, NTT is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton

30    enzyme from <u>Schistosoma japonicum</u>, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from NTT at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-

His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified NTT obtained by these methods can be used directly in the assays shown in Examples XVI and XVII where applicable.

5  **XIII.  Functional Assays**

NTT function is assessed by expressing the sequences encoding NTT at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which

10  contain the cytomegalovirus promoter. 5-10 $\mu$g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 $\mu$g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the

15  recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include

20  changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated

25  Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of NTT on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding NTT and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human

30  immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding NTT and other genes of interest can be analyzed by northern analysis or microarray techniques.

### XIV. Production of NTT Specific Antibodies

NTT substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

5        Alternatively, the NTT amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

10        Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide

15    and anti-NTT activity by, for example, binding the peptide or NTT to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

### XV. Purification of Naturally Occurring NTT Using Specific Antibodies

Naturally occurring or recombinant NTT is substantially purified by immunoaffinity chromatography using antibodies specific for NTT. An immunoaffinity column is constructed by

20    covalently coupling anti-NTT antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing NTT are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of NTT (e.g., high ionic strength buffers in the

25    presence of detergent). The column is eluted under conditions that disrupt antibody/NTT binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and NTT is collected.

### XVI. Identification of Molecules Which Interact with NTT

NTT, or biologically active fragments thereof, are labeled with [125]I Bolton-Hunter reagent.

30    (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled NTT, washed, and any wells with labeled NTT complex are assayed. Data obtained using different concentrations of NTT are used to calculate values for the number, affinity, and association of NTT with the candidate molecules.

Alternatively, molecules interacting with NTT are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

NTT may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT)

5    which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

### XVII.  Demonstration of NTT Activity

Measurements of NTT activity include tracer fluxes and electrophysiological approaches.

10    Tracer fluxes are demonstrated by measuring uptake of labeled substrates into Xenopus laevis oocytes. Oocytes at stages V and VI are injected with NTT mRNA (10 ng per oocyte) and incubated for three days at 18°C in OR2 medium (82.5mM NaCl, 2.5 mM KCl, 1mM $CaCl_2$, 1mM $MgCl_2$, 1mM $Na_2HPO_4$, 5 mM Hepes, 3.8 mM NaOH , 50µg/ml gentamycin, pH 7.8) to allow expression of NTT protein. Oocytes are then transferred to standard uptake medium (100mM NaCl, 2 mM KCl, 1mM

15    $CaCl_2$, 1mM $MgCl_2$, 10 mM Hepes/Tris pH 7.5). Uptake of various neurotransmitters is initiated by adding a $^3H$ substrate to the oocytes. After incubating for 30 minutes, uptake is terminated by washing the oocytes three times in $Na^+$-free medium, measuring the incorporated $^3H$, and comparing with controls. NTT activity is proportional to the level of internalized $^3H$ substrate.

Alternatively, NTT activity can be demonstrated using an electrophysiological assay for ion

20    conductance. Capped NTT mRNA transcribed with T7 polymerase is injected into defolliculated stage V Xenopus oocytes, similar to previously described. Two to seven days later, transport is measured by two-electrode voltage clamp recording. Two-electrode voltage clamp recordings are performed at a holding potential of 50 mV. The data are filtered at 10 Hz and recorded with MacLab digital-to-analog converter and software for data acquisition and analysis (AD Instruments, Castle Hill, Australia). To

25    study the dependence of NTT on external ions, sodium can be replaced by choline or N-methyl-D-glucamine and chloride by gluconate, $NO_3$, or $SO_4$ (Kavanaugh, M.P. et al. (1992) J. Biol. Chem. 267:22007-22009).

In the alternative, choline transporter activity or choline-transporter-like CTL1 protein activity of NTT is determined by measuring choline uptake by yeast transformed with expression

30    vectors harboring polynucleotides encoding NTT. The assay is performed in nitrogen-free medium at 30°C for 10 or 30 min in the presence of 25 nM [$^3H$]choline. The cells are then filtered, and washed. The amount of [$^3H$]choline present in the cells is proportional to the activity of NTT in the cells (O'Regan, S. supra).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments.

5    Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

## Table 1

| Incyte Project ID | Polypeptide SEQ ID NO: | Incyte Polypeptide ID | Polynucleotide SEQ ID NO: | Incyte Polynucleotide ID |
|---|---|---|---|---|
| 7617689 | 1 | 7617689CD1 | 7 | 7617689CB1 |
| 6881669 | 2 | 6881669CD1 | 8 | 6881669CB1 |
| 2742486 | 3 | 2742486CD1 | 9 | 2742486CB1 |
| 71556695 | 4 | 71556695CD1 | 10 | 71556695CB1 |
| 7472800 | 5 | 7472800CD1 | 11 | 7472800CB1 |
| 7475553 | 6 | 7475553CD1 | 12 | 7475553CB1 |

## Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: | Probability score | GenBank Homolog |
|---|---|---|---|---|
| 1 | 7617689CD1 | g202523 | 0 | [Rattus norvegicus] GABA transporter (Borden,L.A. et al. (1992) J. Biol. Chem. 267, 21098-21104) |
| 2 | 6881669CD1 | g298324 | 0 | [Rattus sp.] sodium-dependent neurotransmitter transporter (Uhl,G.R. et al. (1992) Brain Res. Mol. Brain Res. 16 (3-4), 353-359) |
| 3 | 2742486CD1 | g1314290 | 2.10E-19 | [Homo sapiens] vesicular monoamine transporter VMAT1 (Erickson,J.D. et al. (1996) Proc. Natl. Acad. Sci. U.S.A. 93 (10), 5166-5171) |
| 4 | 71556695CD1 | g799339 | 0 | [Homo sapiens] taurine transporter (Miyamoto,Y. et al. (1996) Curr. Eye Res. 15 (3), 345-349) |
| 5 | 7472800CD1 | g181656 | 0 | [Homo sapiens] dopamine transporter (Vandenbergh,D.J. et al. (1992) Brain Res. Mol. Brain Res. 15 (1-2), 161-166) |
| 6 | 7475553CD1 | g914028 | 5.70E-300 | [Rattus sp.] neurotransmitter transporter rB21a (Smith,K.E. et al. (1995) FEBS Lett. 357 (1), 86-92) |

## Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosyla-tion Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 1 | 7617689CD1 | 602 | S10 T596 T525 T230 T555 T587 | N169 N173 N178 N269 | Transmembrane Domains:<br>  N126-W150<br>  S325-Q341<br>  F371-S390<br>  V426-L445<br>  M454-G472 | HMMER |
| | | | | | Sodium:neurotransmitter symporter family (SNF):<br>  R32-P570 | HMMER_PFAM |
| | | | | | Sodium:neurotransmitter symporter family proteins<br>  BL00610:E40-E89, W104-C153, W206-G257, R272-T324, A365-P407, S464-L518, Y524-P546 | BLIMPS_BLOCKS |
| | | | | | Sodium:neurotransmitter symporter family signatures:<br>  N36-T90 | PROFILESCAN |
| | | | | | Sodium/neurotransmitter symporter signature<br>  PR00176:E40-L61, A69-L88, G113-Y139, V233-I250, I315-I335, L369-L388, G453-A473, K493-I513 | BLIMPS_PRINTS |

## Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 1 | | | | | TRANSPORTER NEUROTRANSMITTER SYMPORT GLYCOPROTEIN SODIUM CHLORIDE DEPENDENT GABA PD000448:R32-P570 | BLAST_PRODOM |
| | | | | | TRANSMEMBRANE TRANSPORT PROTEIN TRANSPORTER AMINO ACID PERMEASE GLYCOPROTEIN MEMBRANE PD000214:A258-P546 | BLAST_PRODOM |
| | | | | | SODIUM CHLORIDE DEPENDENT GABA TRANSPORTER NEUROTRANSMITTER GLYCOPROTEIN SYMPORT MULTIGENE PD037836:T555-C602 | BLAST_PRODOM |
| | | | | | SODIUM:NEUROTRANSMITTER SYMPORTER FAMILY DM00572\|P31646\|22-576:E22-R577 DM00572\|P31649\|22-576:E22-R577 DM00572\|P48066\|40-596:K24-R577 DM00572\|P27799\|26-581:E22-Q576 | BLAST_DOMO |
| | | | | | Sodium:neurotransmitter symporter family signature 1: W56-F70 | MOTIFS |
| | | | | | Sodium:neurotransmitter symporter family signature 2: Y139-T159 | MOTIFS |
| 2 | 6881669CD1 | 730 | S129 S137 S19 S217 S25 S250 S425 S55 S699 T191 T287 T399 T489 T49 T522 T656 Y614 | N187 N213 N276 N383 N394 | Transmembrane Domains: L100-L117, V335-K359, S460-V483, K494-F512, M572-V589, L619-R640 | HMMER |
| | | | | | Sodium:neurotransmitter symporter family: R61-R659 | HMMER_PFAM |
| | | | | | Sodium:neurotransmitter symporter family BL00610:Q69-E118, W133-C182, W226-G277, I292-T344, A449-K491, N541-S595, L612-P634 | BLIMPS_BLOCKS |

## Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 2 | | | | | Sodium:neurotransmitter symporter family signatures: N65-L119 | PROFILESCAN |
| | | | | | Sodium/neurotransmitter symporter signature PR00176:Q69-L90, A98-L117, G142-Y168, I253-I270, V335-V355, F453-L472, T530-I550, Y570-V590 | BLIMPS_PRINTS |
| | | | | | TRANSPORTER NEUROTRANSMITTER SYMPORT GLYCOPROTEIN SODIUM CHLORIDE DEPENDENT GABA PD000448:L423-R640, R61-F353 | BLAST_PRODOM |
| | | | | | SODIUM CHLORIDE DEPENDENT TRANSPORTER NEUROTRANSMITTER TRANSPORT NTT4 NTT73 PD037831:N651-D724 | BLAST_PRODOM |
| | | | | | SODIUM AND CHLORIDE DEPENDENT TRANSPORTER NTT73 NEUROTRANSMITTER TRANSPORT SYMPORT PD113985:M1-E60 | BLAST_PRODOM |
| | | | | | SODIUM CHLORIDE DEPENDENT TRANSPORTER NEUROTRANSMITTER SYMPORT NTT4 NTT73 PD037830:K359-E417 | BLAST_PRODOM |
| | | | | | SODIUM:NEUROTRANSMITTER SYMPORTER FAMILY DM00572 Q08469 51-678:E52-G679 DM00572 P31662 50-676:E52-I677 DM00572 S50998 19-616:E56-V636 DM00572 A48716 183-762:S200-N361 | BLAST_DOMO |
| | | | | | Sodium:neurotransmitter symporter family signatures Na_Neurotran_Symp_1 W85-Y99 | MOTIFS |

## Table 3 (cont.).

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 3 | 2742486CD1 | 523 | S503 T52 S88 T420 S510 T511 S9 S49 S362 S498 | | SIMILARITY TO VESICULAR MONOAMINE TRANSPORTERS PD075879: E287-H394 | BLAST-PRODOM |
| | | | | | VESICULAR LUMEN DOMAIN DM03018\|P54219\|172-524: M199-E512 | BLAST-DOMO |
| | | | | | Transmembrane Domains: F168-L187, F265-L283, L399-I418 | HMMER |
| 4 | 71556695CD1 | 649 | S23 T57 S75 T140 T312 S351 S442 T610 S74 T204 S244 T271 S351 T610 | N192 N208 N219 N562 | Sodium:neurotransmitter symporter family BL00610: D78-E127, W142-C191, W247-G298, R313-T365, A406-P448, C505-L559, Y565-P587 | BLIMPS-BLOCKS |
| | | | | | SODIUM:NEUROTRANSMITTER SYMPORTER FAMILY DM00572\|P31649\|22-576: Q69-E613 | BLAST-DOMO |
| | | | | | Sodium:neurotransmitter symporter family neurotransm_transp_1.prf: S74-I128 | PROFILESCAN |
| | | | | | SODIUM/NEUROTRANSMITTER SYMPORTER FAMILY SIGNATURE PR00176: D78-L99, A107-L126, G151-Y177, V274-V291, M356-I376, M410-L429, G494-G514, K534-V554 | BLIMPS-PRINTS |
| | | | | | TRANSPORTER NEUROTRANSMITTER TRANSPORT TRANSMEMBRANE SYMPORT GLYCOPROTEIN SODIUM CHLORIDE DEPENDENT SODIUM DEPENDENT GABA PD000448: R70-P611 | BLAST-PRODOM |
| | | | | | SODIUM:NEUROTRANSMITTER SYMPORTER FAMILY DM00572\|Q00589\|31-588: E60-A618 | BLAST-DOMO |
| | | | | | Sodium:neurotransmitter symporter family signature Na_Neurotran_Symp_1: W94-F108 Na_Neurotran_Symp_2: Y177-T197 | MOTIFS |
| | | | | | Transmembrane Domains: L109-I128, I418-V436, V495-G513, S536-L553 | HMMER |

## Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 4 | | | | | Sodium:neurotransmitter symporter family SNF:<br>R70-P611 | HMMER-PFAM |
| 5 | 7472800CD1 | 625 | S197 S262 S53<br>S583 T476 T485<br>T539 T610 | N181 N188<br>N205 | Transmembrane Domains:<br>L103-F123, I146-T173, A409-A426, F451-<br>N469 | HMMER |
| | | | | | Sodium:neurotransmitter symporter<br>family:<br>R60-G386 W410-P594 | HMMER_PFAM |
| | | | | | Sodium:neurotransmitter symporter family<br>BL00610:D68-E117, V131-C180, W238-<br>G289, R304-T356, P400-Q442, T489-P543,<br>Y548-P570 | BLIMPS_BLOCKS |
| | | | | | Sodium:neurotransmitter symporter family<br>signatures:<br>G64-L118 | PROFILESCAN |
| | | | | | Sodium/neurotransmitter symporter<br>signature<br>PR00176:D68-L89, A97-M116, G140-Y166,<br>V265-L282, I347-F367, S404-I423, R518-<br>V538 | BLIMPS_PRINTS |
| | | | | | TRANSPORTER NEUROTRANSMITTER SYMPORT<br>GLYCOPROTEIN SODIUM CHLORIDE DEPENDENT<br>GABA<br>PD000448:R60-P594 | BLAST_PRODOM |
| | | | | | SODIUM DEPENDENT DOPAMINE TRANSPORTER DA<br>DAT NEUROTRANSMITTER SYMPORT<br>PD023333:M1-D59 | BLAST_PRODOM |
| | | | | | TRANSMEMBRANE TRANSPORT PROTEIN PERMEASE<br>AMINO ACID GLYCOPROTEIN MEMBRANE<br>PD000214:L244-L515, L453-K576, W238-<br>S333, V145-S186 | BLAST_PRODOM |

## Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosyla-tion Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 5 | | | | | SODIUM:NEUROTRANSMITTER SYMPORTER FAMILY<br>DM00572\|P27922\|49-600:N49-D597<br>DM00572\|P51143\|44-598:E56-V601<br>DM00572\|P31652\|69-620:D59-L600<br>DM00572\|P51905\|64-612:V55-E595 | BLAST_DOMO |
| | | | | | Sodium:neurotransmitter symporter family signatures<br>Na_Neurotran_Symp_1 W84-F98<br>Na_Neurotran_Symp_2 Y166-S186 | MOTIFS |
| 6 | 7475553CD1 | 592 | S116 S159 S264<br>S314 S331 S362<br>T133 T192 T229<br>T497 T533 | N131 N357 | Transmembrane Domains:<br>Y83-N103, L174-T199, I277-I297,<br>S392-L415, M508-Y524 | HMMER |
| | | | | | Sodium:neurotransmitter symporter family:<br>R5-T303, C358-Q580 | HMMER-PFAM |
| | | | | | Sodium:neurotransmitter symporter family signatures:<br>A9-L63 | ProfileScan |
| | | | | | Sodium:neurotransmitter:<br>BL00610: W77-C126, W168-G219, Q234-T286, A381-I423, T477-T531, K549-P571, Q13-E62 | BLIMPS-BLOCKS |
| | | | | | Sodium/neurotransmitter:<br>PR00176: Q13-L34, S42-L61, G86-Y112, V195-I212, I277-I297, M385-I404, T466-L486, K506-S526 | BLIMPS-PRINTS |
| | | | | | Sodium:neurotransmitter symporter family:<br>DM00572\|S50998\|19-616: M1-A592<br>DM00572\|Q08469\|51-678: M1-R582<br>DM00572\|P31662\|50-676: R5-L572<br>DM00572\|P28570\|42-603: R5-C139, D136-T303, A367-R582 | BLAST-DOMO |
| | | | | | Neurotransmitter transporter:<br>PD000448: R5-F300, G257-R582 | BLAST-PRODOM |

Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosyla- tion Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 6 | | | | | Orphan neurotransmitter transporter RB21A: PD154252: K301-L360 | BLAST-PRODOM |

## Table 4

| Polynucleotide SEQ ID NO: | Incyte Polynucleotide ID | Sequence Length | Selected Fragment(s) | Sequence Fragments | 5' Position | 3' Position |
|---|---|---|---|---|---|---|
| 7 | 7617689CB1 | 2168 | 1383-1520 | 5407696F6 (BRAMNOT01) | 210 | 742 |
| | 7617689CB1 | 2168 | 1383-1520 | 70501086V1 | 1342 | 1949 |
| | 7617689CB1 | 2168 | 1383-1520 | 5407696F8 (BRAMNOT01) | 204 | 711 |
| | 7617689CB1 | 2168 | 1383-1520 | 70505706V1 | 590 | 1263 |
| | 7617689CB1 | 2168 | 1383-1520 | 757515H1 (BRAITUT02) | 1 | 222 |
| | 7617689CB1 | 2168 | 1383-1520 | 70501894V1 | 741 | 1381 |
| | 7617689CB1 | 2168 | 1383-1520 | 70505220V1 | 1545 | 2168 |
| | 7617689CB1 | 2168 | 1383-1520 | 70501380V1 | 1369 | 1987 |
| 8 | 6881669CB1 | 2709 | 2539-2709, 74-93, 330-610 | 6290242H1 (EPIPUNA01) | 483 | 616 |
| | 6881669CB1 | 2709 | 2539-2709, 74-93, 330-610 | 7186875H1 (BMARTXR02) | 580 | 1179 |
| | 6881669CB1 | 2709 | 2539-2709, 74-93, 330-610 | GBI:g6881669_edit | 113 | 868 |
| | 6881669CB1 | 2709 | 2539-2709, 74-93, 330-610 | 4761061F8 (BRAMNOT01) | 784 | 1392 |
| | 6881669CB1 | 2709 | 2539-2709, 74-93, 330-610 | 7372128H2 (BRAIFEE04) | 1 | 270 |
| | 6881669CB1 | 2709 | 2539-2709, 74-93, 330-610 | 55046086H1 | 1328 | 1953 |
| | 6881669CB1 | 2709 | 2539-2709, 74-93, 330-610 | 6914742F6 (PITUDIR01) | 1836 | 2709 |
| | 6881669CB1 | 2709 | 2539-2709, 74-93, 330-610 | 2289873R6 (BRAINON01) | 1735 | 2174 |
| 9 | 2742486CB1 | 2958 | 1-50, 2756-2816, 498-1495 | 41811117H1 (SINITUT03) | 1 | 216 |
| | 2742486CB1 | 2958 | 1-50, 2756-2816, 498-1495 | g820961 () | 2549 | 2958 |
| | 2742486CB1 | 2958 | 1-50, 2756-2816, 498-1495 | 2896553F6 (KIDNTUT14) | 1109 | 1730 |
| | 2742486CB1 | 2958 | 1-50, 2756-2816, 498-1495 | 7117217H1 (BRAHNOE01) | 749 | 1245 |
| | 2742486CB1 | 2958 | 1-50, 2756-2816, 498-1495 | 2419627R6 (SCORNQON02) | 1599 | 2161 |
| | 2742486CB1 | 2958 | 1-50, 2756-2816, 498-1495 | 2419627T6 (SCORNQON02) | 2289 | 2949 |

## Table 4

| Polynucleotide SEQ ID NO: | Incyte Polynucleotide ID | Sequence Length | Selected Fragment(s) | Sequence Fragments | 5' Position | 3' Position |
|---|---|---|---|---|---|---|
| 7 | 7617689CB1 | 2168 | 1383-1520 | 5407696F6 (BRAMNOT01) | 210 | 742 |
| | 7617689CB1 | 2168 | 1383-1520 | 70501086V1 | 1342 | 1949 |
| | 7617689CB1 | 2168 | 1383-1520 | 5407696F8 (BRAMNOT01) | 204 | 711 |
| | 7617689CB1 | 2168 | 1383-1520 | 70505706V1 | 590 | 1263 |
| | 7617689CB1 | 2168 | 1383-1520 | 757515H1 (BRAITUT02) | 1 | 222 |
| | 7617689CB1 | 2168 | 1383-1520 | 70501894V1 | 741 | 1381 |
| | 7617689CB1 | 2168 | 1383-1520 | 70505220V1 | 1545 | 2168 |
| | 7617689CB1 | 2168 | 1383-1520 | 70501380V1 | 1369 | 1987 |
| 8 | 6881669CB1 | 2709 | 2539-2709, 74-93, 330-610 | 6290242H1 (EPIPUNA01) | 483 | 616 |
| | 6881669CB1 | 2709 | 2539-2709, 74-93, 330-610 | 7186875H1 (BMARTXR02) | 580 | 1179 |
| | 6881669CB1 | 2709 | 2539-2709, 74-93, 330-610 | GBI:g6881669_edit | 113 | 868 |
| | 6881669CB1 | 2709 | 2539-2709, 74-93, 330-610 | 4761061F8 (BRAMNOT01) | 784 | 1392 |
| | 6881669CB1 | 2709 | 2539-2709, 74-93, 330-610 | 7372128H2 (BRAIFEE04) | 1 | 270 |
| | 6881669CB1 | 2709 | 2539-2709, 74-93, 330-610 | 55046086H1 | 1328 | 1953 |
| | 6881669CB1 | 2709 | 2539-2709, 74-93, 330-610 | 6914742F6 (PITUDIR01) | 1836 | 2709 |
| | 6881669CB1 | 2709 | 2539-2709, 74-93, 330-610 | 2289873R6 (BRAINON01) | 1735 | 2174 |
| 9 | 2742486CB1 | 2958 | 1-50, 2756-2816, 498-1495 | 4181117H1 (SINITUT03) | 1 | 216 |
| | 2742486CB1 | 2958 | 1-50, 2756-2816, 498-1495 | g820961 () | 2549 | 2958 |
| | 2742486CB1 | 2958 | 1-50, 2756-2816, 498-1495 | 2896553F6 (KIDNTUT14) | 1109 | 1730 |
| | 2742486CB1 | 2958 | 1-50, 2756-2816, 498-1495 | 7117217H1 (BRAHNOE01) | 749 | 1245 |
| | 2742486CB1 | 2958 | 1-50, 2756-2816, 498-1495 | 2419627R6 (SCORNON02) | 1599 | 2161 |
| | 2742486CB1 | 2958 | 1-50, 2756-2816, 498-1495 | 2419627T6 (SCORNON02) | 2289 | 2949 |

| Polynucleotide SEQ ID NO: | Incyte Polynucleotide ID | Sequence Length | Selected Fragment(s) | Sequence Fragments | 5' Position | 3' Position |
|---|---|---|---|---|---|---|
| | 2742486CB1 | 2958 | 1-50, 2756-2816, 498-1495 | 55047291H1 | 533 | 839 |
| | 2742486CB1 | 2958 | 1-50, 2756-2816, 498-1495 | 5777360H1 (BRAINOT20) | 2397 | 2950 |
| | 2742486CB1 | 2958 | 1-50, 2756-2816, 498-1495 | 55047293J1 | 163 | 836 |
| | 2742486CB1 | 2958 | 1-50, 2756-2816, 498-1495 | 1398158F6 (BRAITUT08) | 1791 | 2356 |
| 10 | 71556695CB1 | 2135 | 1-261, 2048-2135 | 7685653H1 (BRABDIK02) | 1526 | 2135 |
| | 71556695CB1 | 2135 | 1-261, 2048-2135 | 7723816H1 (THYRDIE01) | 1142 | 1484 |
| | 71556695CB1 | 2135 | 1-261, 2048-2135 | g799338_CD | 273 | 2135 |
| | 71556695CB1 | 2135 | 1-261, 2048-2135 | 7641640H1 (SEMVTDE01) | 504 | 1178 |
| | 71556695CB1 | 2135 | 1-261, 2048-2135 | 7620645H1 (HEARFEE03) | 1 | 587 |
| 11 | 7472800CB1 | 1997 | 1339-1469, 1-116, 1576-1607, 893-1145, 1950-1997, 689-744 | 6866221H1 (BRAGNON02) | 1 | 698 |
| | 7472800CB1 | 1997 | 1339-1469, 1-116, 1576-1607, 893-1145, 1950-1997, 689-744 | FL7472800_g7108462_g181656 | 120 | 1997 |
| 12 | 7475553CB1 | 2774 | 1-259, 2047-2318, 2715-2774 | FL7475553_g3702356_g914028 | 1 | 1812 |
| | 7475553CB1 | 2774 | 1-259, 2047-2318, 2715-2774 | 70646995V1 | 2091 | 2774 |
| | 7475553CB1 | 2774 | 1-259, 2047-2318, 2715-2774 | 71867729V1 | 1844 | 2494 |
| | 7475553CB1 | 2774 | 1-259, 2047-2318, 2715-2774 | 71869071V1 | 1790 | 2441 |

## Table 5

| Polynucleotide SEQ ID NO: | Incyte Project ID | Representative Library |
|---|---|---|
| 7 | 7617689CB1 | BRAITUT02 |
| 8 | 6881669CB1 | BRABDIR02 |
| 9 | 2742486CB1 | BRSTTUT14 |
| 10 | 71556695CB1 | DRGTNOT01 |
| 11 | 7472800CB1 | BRAGNON02 |
| 12 | 7475553CB1 | BRAINOT09 |

## Table 6

| Library | Vector | Library Description |
|---|---|---|
| BRAGNON02 | pINCY | This normalized substantia nigra tissue library was constructed from 4.2 10e7 independent clones from a substantia nigra tissue library. Starting RNA was made from RNA isolated from substantia nigra tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis. Patient history included sepsis, cholangitis, post-operative atelectasis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolonephrosclerosis, nodular colloidal goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used. |
| BRABDIR02 | pINCY | Library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Serologies were negative. Patient history included Huntington's disease, emphysema, and tobacco abuse (3-4 packs per day for 40 years). |
| BRAINOT09 | pINCY | Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus, who died at 23 weeks' gestation. |
| BRAITUT02 | PSPORT1 | Library was constructed using RNA isolated from brain tumor tissue removed from the frontal lobe of a 58-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated a grade 2 metastatic hypernephroma. Patient history included a grade 2 renal cell carcinoma, insomnia, and chronic airway obstruction. Family history included a malignant neoplasm of the kidney. |

## Table 6 (cont.)

| Library | Vector | Library Description |
|---|---|---|
| BRSTTUT14 | pINCY | Library was constructed using RNA isolated from breast tumor tissue removed from a 62-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated an invasive grade 3 (of 4), nuclear grade 3 (of 3) adenocarcinoma, ductal type. Ductal carcinoma in situ, comedo type, comprised 60% of the tumor mass. Metastatic adenocarcinoma was identified in one (of 14) axillary lymph nodes with no perinodal extension. Tumor cells were strongly positive for estrogen receptors and weakly positive for progesterone receptors. Patient history included benign colon neoplasm, hyperlipidemia, cardiac dysrhythmia, and obesity. Family history included atherosclerotic coronary artery disease, myocardial infarction, colon cancer, ovarian cancer, lung cancer, and cerebrovascular disease. |
| DRGTNOT01 | pINCY | Library was constructed using RNA isolated from dorsal root ganglion tissue removed from the thoracic spine of a 32-year-old Caucasian male who died from acute pulmonary edema and bronchopneumonia, bilateral pleural and pericardial effusions, and malignant lymphoma (natural killer cell type). Patient history included probable cytomegalovirus, infection, hepatic congestion and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, and Bell's palsy. Surgeries included colonoscopy, large intestine biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy; treatment included radiation therapy. |

## Table 7

| Program | Description | Reference | Parameter Threshold |
|---|---|---|---|
| ABI FACTURA | A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences. | Applied Biosystems, Foster City, CA. | Mismatch <50% |
| ABI/PARACEL FDF | A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences. | Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA. | |
| ABI AutoAssembler | A program that assembles nucleic acid sequences. | Applied Biosystems, Foster City, CA. | |
| BLAST | A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx. | Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402. | *ESTs*: Probability value= 1.0E-8 or less $$ $$ *Full Length sequences*: Probability value= 1.0E-10 or less |
| FASTA | A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch. | Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489. | *ESTs*: fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences*: fastx score=100 or greater |
| BLIMPS | A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions. | Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424. | Probability value= 1.0E-3 or less |
| HMMER | An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM. | Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350. | *PFAM hits*: Probability value= 1.0E-3 or less *Signal peptide hits*: Score= 0 or greater |

## Table 7 (cont.)

| Program | Description | Reference | Parameter Threshold |
|---|---|---|---|
| ProfileScan | An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite. | Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221. | Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1. |
| Phred | A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability. | Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194. | |
| Phrap | A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences. | Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA. | Score= 120 or greater; Match length= 56 or greater |
| Consed | A graphical tool for viewing and editing Phrap assemblies. | Gordon, D. et al. (1998) Genome Res. 8:195-202. | |
| SPScan | A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides. | Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439. | Score=3.5 or greater |
| TMAP | A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation. | Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371. | |
| TMHMMER | A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation. | Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182. | |
| Motifs | A program that searches amino acid sequences for patterns that matched those defined in Prosite. | Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI | |

What is claimed is:

     1. An isolated polypeptide selected from the group consisting of:

     a) a polypeptide comprising an amino acid sequence selected from the group consisting of

5   SEQ ID NO:1-6,

     b) a naturally occurring polypeptide comprising an amino acid sequence at least 90%

identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6,

     c) a biologically active fragment of a polypeptide having an amino acid sequence selected

from the group consisting of SEQ ID NO:1-6, and

10      d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from

the group consisting of SEQ ID NO:1-6.


     2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-6.


15      3. An isolated polynucleotide encoding a polypeptide of claim 1.


     4. An isolated polynucleotide encoding a polypeptide of claim 2.


     5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID

20   NO:7-12.


     6. A recombinant polynucleotide comprising a promoter sequence operably linked to a

polynucleotide of claim 3.


25      7. A cell transformed with a recombinant polynucleotide of claim 6.


     8. A transgenic organism comprising a recombinant polynucleotide of claim 6.


     9. A method for producing a polypeptide of claim 1, the method comprising:

30      a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said

cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide

comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim

1, and

     b) recovering the polypeptide so expressed.

35

10. An isolated antibody which specifically binds to a polypeptide of claim 1.


11. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting

5  of SEQ ID NO:7-12,

b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90%

identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

10        e) an RNA equivalent of a)-d).


12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a

polynucleotide of claim 11.


15        13. A method for detecting a target polynucleotide in a sample, said target polynucleotide

having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides

comprising a sequence complementary to said target polynucleotide in the sample, and which probe

specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization

20  complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if

present, the amount thereof.


14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

25

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide

having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction

amplification, and

30        b) detecting the presence or absence of said amplified target polynucleotide or fragment

thereof, and, optionally, if present, the amount thereof.


16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable

excipient.

35

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

18. A method for treating a disease or condition associated with decreased expression of
5    functional NTT, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:
10       a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
         b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.
15

21. A method for treating a disease or condition associated with decreased expression of functional NTT, comprising administering to a patient in need of such treatment a composition of claim 20.

20       22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:
         a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
         b) detecting antagonist activity in the sample.

25       23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional NTT, comprising administering to a patient in need of such treatment a composition of claim 23.
30
         25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:
         a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

29. A diagnostic test for a condition or disease associated with the expression of NTT in a biological sample comprising the steps of:

    a) combining the biological sample with an antibody of claim 10, under conditions suitable
5   for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and

    b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

10      a) a chimeric antibody,

        b) a single chain antibody,

        c) a Fab fragment,

        d) a $F(ab')_2$ fragment, or

        e) a humanized antibody.

15

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of NTT in a subject, comprising administering to said subject an effective amount of the composition of claim
20  31.

33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of NTT in
25  a subject, comprising administering to said subject an effective amount of the composition of claim
    33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim
    10 comprising:

30      a) immunizing an animal with a polypeptide having an amino acid sequence selected from
    the group consisting of SEQ ID NO:1-6, or an immunogenic fragment thereof, under conditions to
    elicit an antibody response;

        b) isolating antibodies from said animal; and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

5          36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim
10    10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

b) isolating antibody producing cells from the animal;

15          c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;

d) culturing the hybridoma cells; and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

20

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

25          41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

30

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6 in a sample, comprising the steps of:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6 in the sample.

5      44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6 from a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) separating the antibody from the sample and obtaining the purified polypeptide having an
10    amino acid sequence selected from the group consisting of SEQ ID NO:1-6. ·

45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

15

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.
20

50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

51. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:7.
25

52. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:8.

53. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:9.

30    54. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:10.

55. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:11.

56. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:12.

```
<110> INCYTE GENOMICS, INC.
      SANJANWALA, Madhu S.
      WALIA, Narinder K.
      TRIBOULEY, Catherine M.
      YUE, Henry
      GANDHI, Ameena R.
      DING, Li
      YAO, Monique G.
      LAL, Preeti
      BAUGHN, Mariah R.
      HAFALIA, April
      ELLIOTT, Vicki S.
      PATTERSON, Chandra
      RAMKUMAR, Jaya

<120> NEUROTRANSMITTER TRANSPORTERS

<130> PI-0101 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/205,518; 60/213,956; 60/215,105; 60/218,947; 60/218,947
<151> 2000-05-19; 2000-06-22; 2000-06-28; 2000-07-14; 2000-07-14

<160> 12
<170> PERL Program

<210> 1
<211> 602
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7617689CD1

<400> 1
Met Asp Ser Arg Val Ser Gly Thr Thr Ser Asn Gly Glu Thr Lys
 1               5                  10                  15
Pro Val Tyr Pro Val Met Glu Lys Lys Glu Glu Asp Gly Thr Leu
                20                  25                  30
Glu Arg Gly His Trp Asn Asn Lys Met Glu Phe Val Leu Ser Val
                35                  40                  45
Ala Gly Glu Ile Ile Gly Leu Gly Asn Val Trp Arg Phe Pro Tyr
                50                  55                  60
Leu Cys Tyr Lys Asn Gly Gly Gly Ala Phe Phe Ile Pro Tyr Leu
                65                  70                  75
Val Phe Leu Phe Thr Cys Gly Ile Pro Val Phe Leu Leu Glu Thr
                80                  85                  90
Ala Leu Gly Gln Tyr Thr Ser Gln Gly Gly Val Thr Ala Trp Arg
                95                 100                 105
Lys Ile Cys Pro Ile Phe Glu Gly Ile Gly Tyr Ala Ser Gln Met
               110                 115                 120
Ile Val Ile Leu Leu Asn Val Tyr Tyr Ile Ile Val Leu Ala Trp
               125                 130                 135
Ala Leu Phe Tyr Leu Phe Ser Ser Phe Thr Ile Asp Leu Pro Trp
               140                 145                 150
Gly Gly Cys Tyr His Glu Trp Asn Thr Glu His Cys Met Glu Phe
               155                 160                 165
Gln Lys Thr Asn Gly Ser Leu Asn Gly Thr Ser Glu Asn Ala Thr
               170                 175                 180
Ser Pro Val Ile Glu Phe Trp Glu Arg Arg Val Leu Lys Ile Ser
               185                 190                 195
Asp Gly Ile Gln His Leu Gly Ala Leu Arg Trp Glu Leu Ala Leu
               200                 205                 210
Cys Leu Leu Leu Ala Trp Val Ile Cys Tyr Phe Cys Ile Trp Lys
```

```
                    215                 220             225
Gly Val Lys Ser Thr Gly Lys Val Val Tyr Phe Thr Ala Thr Phe
                    230                 235             240
Pro Tyr Leu Met Leu Val Val Leu Leu Ile Arg Gly Val Thr Leu
                    245                 250             255
Pro Gly Ala Ala Gln Gly Ile Gln Phe Tyr Leu Tyr Pro Asn Leu
                    260                 265             270
Thr Arg Leu Trp Asp Pro Gln Val Trp Met Asp Ala Gly Thr Gln
                    275                 280             285
Ile Phe Phe Ser Phe Ala Ile Cys Leu Gly Cys Leu Thr Ala Leu
                    290                 295             300
Gly Ser Tyr Asn Lys Tyr His Asn Asn Cys Tyr Arg Asp Cys Ile
                    305                 310             315
Ala Leu Cys Phe Leu Asn Ser Gly Thr Ser Phe Val Ala Gly Phe
                    320                 325             330
Ala Ile Phe Ser Ile Leu Gly Phe Met Ser Gln Glu Gln Gly Val
                    335                 340             345
Pro Ile Ser Glu Val Ala Glu Ser Gly Pro Gly Leu Ala Phe Ile
                    350                 355             360
Ala Tyr Pro Arg Ala Val Val Met Leu Pro Phe Ser Pro Leu Trp
                    365                 370             375
Ala Cys Cys Phe Phe Phe Met Val Val Leu Leu Gly Leu Asp Ser
                    380                 385             390
Gln Phe Val Cys Val Glu Ser Leu Val Thr Ala Leu Val Asp Met
                    395                 400             405
Tyr Pro His Val Phe Arg Lys Lys Asn Arg Arg Glu Val Leu Ile
                    410                 415             420
Leu Gly Val Ser Val Val Ser Phe Leu Val Gly Leu Ile Met Leu
                    425                 430             435
Thr Glu Gly Gly Met Tyr Val Phe Gln Leu Phe Asp Tyr Tyr Ala
                    440                 445             450
Ala Ser Gly Met Cys Leu Leu Phe Val Ala Ile Phe Glu Ser Leu
                    455                 460             465
Cys Val Ala Trp Val Tyr Gly Ala Lys Arg Phe Tyr Asp Asn Ile
                    470                 475             480
Glu Asp Met Ile Gly Tyr Arg Pro Trp Pro Leu Ile Lys Tyr Cys
                    485                 490             495
Trp Leu Phe Leu Thr Pro Ala Val Cys Thr Ala Thr Phe Leu Phe
                    500                 505             510
Ser Leu Ile Lys Tyr Thr Pro Leu Thr Tyr Asn Lys Lys Tyr Thr
                    515                 520             525
Tyr Pro Trp Trp Gly Asp Ala Leu Gly Trp Leu Leu Ala Leu Ser
                    530                 535             540
Ser Met Val Cys Ile Pro Ala Trp Ser Leu Tyr Arg Leu Gly Thr
                    545                 550             555
Leu Lys Gly Pro Phe Arg Glu Arg Ile Arg Gln Leu Met Cys Pro
                    560                 565             570
Ala Glu Asp Leu Pro Gln Arg Asn Pro Ala Gly Pro Ser Ala Pro
                    575                 580             585
Ala Thr Pro Arg Thr Ser Leu Leu Arg Leu Thr Glu Leu Glu Ser
                    590                 595             600
His Cys
```

<210> 2
<211> 730
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6881669CD1

<400> 2
```
Met Pro Lys Asn Ser Lys Val Val Lys Arg Glu Leu Asp Asp Asp
 1               5                  10              15
Val Thr Glu Ser Val Lys Asp Leu Leu Ser Asn Glu Asp Ala Ala
```

```
                    20                          25                          30
Asp Asp Ala Phe Lys Thr Ser Glu Leu Ile Val Asp Gly Gln Glu
                    35                          40                          45
Glu Lys Asp Thr Asp Val Glu Glu Gly Ser Glu Val Glu Asp Glu
                    50                          55                          60
Arg Pro Ala Trp Asn Ser Lys Leu Gln Tyr Ile Leu Ala Gln Val
                    65                          70                          75
Gly Phe Ser Val Gly Leu Gly Asn Val Trp Arg Phe Pro Tyr Leu
                    80                          85                          90
Cys Gln Lys Asn Gly Gly Gly Ala Tyr Leu Leu Pro Tyr Leu Ile
                    95                         100                         105
Leu Leu Met Val Ile Gly Ile Pro Leu Phe Phe Leu Glu Leu Ser
                   110                         115                         120
Val Gly Gln Arg Ile Arg Arg Gly Ser Ile Gly Val Trp Asn Tyr
                   125                         130                         135
Ile Ser Pro Lys Leu Gly Gly Ile Gly Phe Ala Ser Cys Val Val
                   140                         145                         150
Cys Tyr Phe Val Ala Leu Tyr Tyr Asn Val Ile Ile Gly Trp Ser
                   155                         160                         165
Leu Phe Tyr Phe Ser Gln Ser Phe Gln Gln Pro Leu Pro Trp Asp
                   170                         175                         180
Gln Cys Pro Leu Val Lys Asn Ala Ser His Thr Phe Val Glu Pro
                   185                         190                         195
Glu Cys Glu Gln Ser Ser Ala Thr Thr Tyr Tyr Trp Tyr Arg Glu
                   200                         205                         210
Ala Leu Asn Ile Ser Ser Ser Ile Ser Glu Ser Gly Gly Leu Asn
                   215                         220                         225
Trp Lys Met Thr Ile Cys Leu Leu Ala Ala Trp Val Met Val Cys
                   230                         235                         240
Leu Ala Met Ile Lys Gly Ile Gln Ser Ser Gly Lys Ile Ile Tyr
                   245                         250                         255
Phe Ser Ser Leu Phe Pro Tyr Val Val Leu Ile Cys Phe Leu Ile
                   260                         265                         270
Arg Ala Phe Leu Leu Asn Gly Ser Ile Asp Gly Ile Arg His Met
                   275                         280                         285
Phe Thr Pro Lys Leu Glu Ile Met Leu Glu Pro Lys Val Trp Arg
                   290                         295                         300
Glu Ala Ala Thr Gln Val Phe Phe Ala Leu Gly Leu Gly Phe Gly
                   305                         310                         315
Gly Val Ile Ala Phe Ser Ser Tyr Asn Lys Arg Asp Asn Asn Cys
                   320                         325                         330
His Phe Asp Ala Val Leu Val Ser Phe Ile Asn Phe Phe Thr Ser
                   335                         340                         345
Val Leu Ala Thr Leu Val Val Phe Ala Val Leu Gly Phe Lys Ala
                   350                         355                         360
Asn Val Ile Asn Glu Lys Cys Ile Thr Gln Asn Ser Glu Thr Ile
                   365                         370                         375
Met Lys Phe Leu Lys Met Gly Asn Ile Ser Gln Asp Ile Ile Pro
                   380                         385                         390
His His Ile Asn Leu Ser Thr Val Thr Ala Glu Asp Tyr His Leu
                   395                         400                         405
Val Tyr Asp Ile Ile Gln Lys Val Lys Glu Glu Glu Phe Pro Ala
                   410                         415                         420
Leu His Leu Asn Ser Cys Lys Ile Glu Glu Glu Leu Asn Lys Ala
                   425                         430                         435
Val Gln Gly Thr Gly Leu Ala Phe Ile Ala Phe Thr Glu Ala Met
                   440                         445                         450
Thr His Phe Pro Ala Ser Pro Phe Trp Ser Val Met Phe Phe Leu
                   455                         460                         465
Met Leu Val Asn Leu Gly Leu Gly Ser Met Phe Gly Thr Ile Glu
                   470                         475                         480
Gly Ile Val Thr Pro Ile Val Asp Thr Phe Lys Val Arg Lys Glu
                   485                         490                         495
Ile Leu Thr Val Ile Cys Cys Leu Leu Ala Phe Cys Ile Gly Leu
                   500                         505                         510
Ile Phe Val Gln Arg Ser Gly Asn Tyr Phe Val Thr Met Phe Asp
                   515                         520                         525
```

3/14

```
Asp Tyr Ser Ala Thr Leu Pro Leu Leu Ile Val Val Ile Leu Glu
                530                 535                 540
Asn Ile Ala Val Cys Phe Val Tyr Gly Ile Asp Lys Phe Met Glu
                545                 550                 555
Asp Leu Lys Asp Met Leu Gly Phe Ala Pro Ser Arg Tyr Tyr Tyr
                560                 565                 570
Tyr Met Trp Lys Tyr Ile Ser Pro Leu Met Leu Leu Ser Leu Leu
                575                 580                 585
Ile Ala Ser Val Val Asn Met Gly Leu Ser Pro Pro Gly Tyr Asn
                590                 595                 600
Ala Trp Ile Glu Asp Lys Ala Ser Glu Glu Phe Leu Ser Tyr Pro
                605                 610                 615
Thr Trp Gly Leu Val Val Cys Val Ser Leu Val Val Phe Ala Ile
                620                 625                 630
Leu Pro Val Pro Val Val Phe Ile Val Arg Arg Phe Asn Leu Ile
                635                 640                 645
Asp Asp Ser Ser Gly Asn Leu Ala Ser Val Thr Tyr Lys Arg Gly
                650                 655                 660
Arg Val Leu Lys Glu Pro Val Asn Leu Glu Gly Asp Asp Thr Ser
                665                 670                 675
Leu Ile His Gly Lys Ile Pro Ser Glu Met Pro Ser Pro Asn Phe
                680                 685                 690
Gly Lys Asn Ile Tyr Arg Lys Gln Ser Gly Ser Pro Thr Leu Asp
                695                 700                 705
Thr Ala Pro Asn Gly Arg Tyr Gly Ile Gly Tyr Leu Met Ala Asp
                710                 715                 720
Ile Met Pro Asp Met Pro Glu Ser Asp Leu
                725                 730
```

<210> 3
<211> 523
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2742486CD1

<400> 3
```
Met Gln Asp Cys Gly Asp Pro Glu Ser Trp Lys Ala Leu Gly Ser
  1               5                  10                  15
Ser Gly Pro Gly Gly Ser Leu Pro Ala Gly Ser Arg Gly Cys Gln
                 20                  25                  30
Val Pro Ala Glu Gln Arg Pro Pro Ala Arg Val Pro Val Ala Pro
                 35                  40                  45
Trp Arg Ser Ser Pro Arg Thr Trp Leu Glu Ala Gly Ser Trp Gly
                 50                  55                  60
Ala Gly Val His Ala Pro Gly Met Glu Ala Leu Gly Asp Leu Glu
                 65                  70                  75
Gly Pro Arg Ala Pro Gly Gly Asp Asp Pro Ala Gly Ser Ala Gly
                 80                  85                  90
Glu Thr Pro Gly Trp Leu Ser Arg Glu Gln Val Phe Val Leu Ile
                 95                 100                 105
Ser Ala Ala Ser Val Asn Leu Gly Ser Met Met Cys Tyr Ser Ile
                110                 115                 120
Leu Gly Pro Phe Phe Pro Lys Glu Ala Glu Lys Lys Gly Ala Ser
                125                 130                 135
Asn Thr Ile Ile Gly Met Ile Phe Gly Cys Phe Ala Leu Phe Glu
                140                 145                 150
Leu Leu Ala Ser Leu Val Phe Gly Asn Tyr Leu Val His Ile Gly
                155                 160                 165
Ala Lys Phe Met Phe Val Ala Gly Met Phe Val Ser Gly Gly Val
                170                 175                 180
Thr Ile Leu Phe Gly Val Leu Asp Arg Val Pro Asp Gly Pro Val
                185                 190                 195
Phe Ile Ala Met Cys Phe Leu Val Arg Val Met Asp Ala Val Ser
                200                 205                 210
```

                                4/14

```
Phe Ala Ala Ala Met Thr Ala Ser Ser Ser Ile Leu Ala Lys Ala
            215                 220                 225
Phe Pro Asn Asn Val Ala Thr Val Leu Gly Ser Leu Glu Thr Phe
            230                 235                 240
Ser Gly Leu Gly Leu Ile Leu Gly Pro Pro Val Gly Gly Phe Leu
            245                 250                 255
Tyr Gln Ser Phe Gly Tyr Glu Val Pro Phe Ile Val Leu Gly Cys
            260                 265                 270
Val Val Leu Leu Met Val Pro Leu Asn Met Tyr Ile Leu Pro Asn
            275                 280                 285
Tyr Glu Ser Asp Pro Gly Glu His Ser Phe Trp Lys Leu Ile Ala
            290                 295                 300
Leu Pro Lys Val Gly Leu Ile Ala Phe Val Ile Asn Ser Leu Ser
            305                 310                 315
Ser Cys Phe Gly Phe Leu Asp Pro Thr Leu Ser Leu Phe Val Leu
            320                 325                 330
Glu Lys Phe Asn Leu Pro Ala Gly Tyr Val Gly Leu Val Phe Leu
            335                 340                 345
Gly Met Ala Leu Ser Tyr Ala Ile Ser Ser Pro Leu Phe Gly Leu
            350                 355                 360
Leu Ser Asp Lys Arg Pro Pro Leu Arg Lys Trp Leu Leu Val Phe
            365                 370                 375
Gly Asn Leu Ile Thr Ala Gly Cys Tyr Met Leu Leu Gly Pro Val
            380                 385                 390
Pro Ile Leu His Ile Lys Ser Gln Leu Trp Leu Leu Val Leu Ile
            395                 400                 405
Leu Val Val Ser Gly Leu Ser Ala Gly Met Ser Ile Ile Pro Thr
            410                 415                 420
Phe Pro Glu Ile Leu Ser Cys Ala His Glu Asn Gly Phe Glu Glu
            425                 430                 435
Gly Leu Ser Thr Leu Gly Leu Val Ser Gly Leu Phe Ser Ala Met
            440                 445                 450
Gly Ser Ile Gly Ala Phe Met Gly Pro Thr Leu Gly Gly Phe Leu
            455                 460                 465
Tyr Glu Lys Ile Gly Phe Glu Trp Ala Ala Ala Ile Gln Gly Leu
            470                 475                 480
Trp Ala Leu Ile Ser Gly Leu Ala Met Gly Leu Phe Tyr Leu Leu
            485                 490                 495
Glu Tyr Ser Arg Arg Lys Arg Ser Lys Ser Gln Asn Ile Leu Ser
            500                 505                 510
Thr Glu Glu Glu Arg Thr Thr Leu Leu Pro Asn Glu Thr
            515                 520
```

<210> 4
<211> 649
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71556695CD1

<400> 4
```
Met Cys Ala Glu Ser Ala Asp Gly Ala Ala Leu Gly Pro Ala Ser
 1               5                  10                  15
Gly Cys Gly Asn Arg Pro Pro Ser Thr Gln Glu Ser Lys Glu Met
            20                  25                  30
Ala Thr Lys Glu Lys Leu Gln Cys Leu Lys Asp Phe His Lys Asp
            35                  40                  45
Ile Leu Lys Pro Ser Pro Gly Lys Ser Pro Gly Thr Arg Pro Glu
            50                  55                  60
Asp Glu Ala Glu Gly Lys Pro Pro Gln Arg Glu Lys Trp Ser Ser
            65                  70                  75
Lys Ile Asp Phe Val Leu Ser Val Ala Gly Gly Phe Val Gly Leu
            80                  85                  90
Gly Asn Val Trp Arg Phe Pro Tyr Leu Cys Tyr Lys Asn Gly Gly
            95                  100                 105
```

```
Gly Ala Phe Leu Ile Pro Tyr Phe Ile Phe Leu Phe Gly Ser Gly
            110                 115                 120
Leu Pro Val Phe Phe Leu Glu Ile Ile Ile Gly Gln Tyr Thr Ser
            125                 130                 135
Glu Gly Gly Ile Thr Cys Trp Glu Lys Ile Cys Pro Leu Phe Ser
            140                 145                 150
Gly Ile Gly Tyr Ala Ser Val Val Ile Val Ser Leu Leu Asn Val
            155                 160                 165
Tyr Tyr Ile Val Ile Leu Ala Trp Ala Thr Tyr Tyr Leu Phe Gln
            170                 175                 180
Ser Phe Gln Lys Glu Leu Pro Trp Ala His Cys Asn His Ser Trp
            185                 190                 195
Asn Thr Pro His Cys Met Glu Asp Thr Met Arg Lys Asn Lys Ser
            200                 205                 210
Val Trp Ile Thr Ile Ser Ser Thr Asn Phe Thr Ser Pro Val Ile
            215                 220                 225
Glu Phe Trp Glu Arg Asn Val Leu Ser Leu Ser Pro Gly Ile Asp
            230                 235                 240
His Pro Gly Ser Leu Lys Trp Asp Leu Ala Leu Cys Leu Leu Leu
            245                 250                 255
Val Trp Leu Val Cys Phe Phe Cys Ile Trp Lys Gly Val Arg Ser
            260                 265                 270
Thr Gly Lys Val Val Tyr Phe Thr Ala Thr Phe Pro Phe Ala Met
            275                 280                 285
Leu Leu Val Leu Leu Val Arg Gly Leu Thr Leu Pro Gly Ala Gly
            290                 295                 300
Ala Gly Ile Lys Phe Tyr Leu Tyr Pro Asp Ile Thr Arg Leu Glu
            305                 310                 315
Asp Pro Gln Val Trp Ile Asp Ala Gly Thr Gln Ile Phe Phe Ser
            320                 325                 330
Tyr Ala Ile Cys Leu Gly Ala Met Thr Ser Leu Gly Ser Tyr Asn
            335                 340                 345
Lys Tyr Lys Tyr Asn Ser Tyr Arg Asp Cys Met Leu Leu Gly Cys
            350                 355                 360
Leu Asn Ser Gly Thr Ser Phe Val Ser Gly Phe Ala Ile Phe Ser
            365                 370                 375
Ile Leu Gly Phe Met Ala Gln Glu Gln Gly Val Asp Ile Ala Asp
            380                 385                 390
Val Ala Glu Ser Gly Pro Gly Leu Ala Phe Ile Ala Tyr Pro Lys
            395                 400                 405
Ala Val Thr Met Met Pro Leu Pro Thr Phe Trp Ser Ile Leu Phe
            410                 415                 420
Phe Ile Met Leu Leu Leu Leu Gly Leu Asp Ser Gln Phe Val Glu
            425                 430                 435
Val Glu Gly Gln Ile Thr Ser Leu Val Asp Leu Tyr Pro Ser Phe
            440                 445                 450
Leu Arg Lys Gly Tyr Arg Arg Glu Ile Phe Ile Ala Phe Val Cys
            455                 460                 465
Ser Ile Ser Tyr Leu Leu Gly Leu Thr Met Val Thr Glu Gly Gly
            470                 475                 480
Met Tyr Val Phe Gln Leu Phe Asp Tyr Tyr Ala Ala Ser Gly Val
            485                 490                 495
Cys Leu Leu Trp Val Ala Phe Phe Glu Cys Phe Val Ile Ala Trp
            500                 505                 510
Ile Tyr Gly Gly Asp Asn Leu Tyr Asp Gly Ile Glu Asp Met Ile
            515                 520                 525
Gly Tyr Arg Pro Gly Pro Trp Met Lys Tyr Ser Trp Ala Val Ile
            530                 535                 540
Thr Pro Val Leu Cys Val Gly Cys Phe Ile Phe Ser Leu Val Lys
            545                 550                 555
Tyr Val Pro Leu Thr Tyr Asn Lys Thr Tyr Val Tyr Pro Asn Trp
            560                 565                 570
Ala Ile Gly Leu Gly Trp Ser Leu Ala Leu Ser Ser Met Leu Cys
            575                 580                 585
Val Pro Leu Val Ile Val Ile Arg Leu Cys Gln Thr Glu Gly Pro
            590                 595                 600
Phe Leu Val Arg Val Lys Tyr Leu Leu Thr Pro Arg Glu Pro Asn
```

```
                    605                    610                    615
Arg Trp Ala Val Glu Arg Glu Gly Ala Thr Pro Tyr Asn Ser Arg
                    620                    625                    630
Thr Val Met Asn Gly Ala Leu Val Lys Pro Thr His Ile Ile Val
                    635                    640                    645
Glu Thr Met Met
```

<210> 5.
<211> 625
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472800CD1

<220>
<221> unsure
<222> 564
<223> unknown or other

```
<400> 5
Met Ser Lys Ser Lys Cys Ser Val Gly Leu Met Ser Ser Val Val
 1               5                    10                    15
Ala Pro Ala Lys Glu Pro Asn Ala Val Gly Pro Lys Glu Val Glu
                20                    25                    30
Leu Ile Leu Val Lys Glu Gln Asn Gly Val Gln Leu Thr Ser Ser
                35                    40                    45
Thr Leu Thr Asn Pro Arg Gln Ser Pro Val Glu Ala Gln Asp Arg
                50                    55                    60
Glu Thr Trp Gly Lys Lys Ile Asp Phe Leu Leu Ser Val Ile Gly
                65                    70                    75
Phe Ala Val Asp Leu Ala Asn Val Trp Arg Phe Pro Tyr Leu Cys
                80                    85                    90
Tyr Lys Asn Gly Gly Gly Ala Phe Leu Val Pro Tyr Leu Leu Phe
                95                    100                   105
Met Val Ile Ala Gly Met Pro Leu Phe Tyr Met Glu Leu Ala Leu
                110                   115                   120
Gly Gln Phe Asn Arg Glu Gly Ala Ala Gly Val Trp Lys Ile Cys
                125                   130                   135
Pro Ile Leu Lys Gly Val Gly Phe Thr Val Ile Leu Ile Ser Leu
                140                   145                   150
Tyr Val Gly Phe Phe Tyr Asn Val Ile Ile Ala Trp Ala Leu His
                155                   160                   165
Tyr Leu Phe Ser Ser Phe Thr Thr Glu Leu Pro Trp Ile His Cys
                170                   175                   180
Asn Asn Ser Trp Asn Ser Pro Asn Cys Ser Asp Ala His Pro Gly
                185                   190                   195
Asp Ser Ser Gly Asp Ser Ser Gly Leu Asn Asp Thr Phe Gly Thr
                200                   205                   210
Thr Pro Ala Ala Glu Tyr Phe Glu Arg Gly Val Leu His Leu His
                215                   220                   225
Gln Ser His Gly Ile Asp Asp Leu Gly Pro Pro Arg Trp Gln Leu
                230                   235                   240
Thr Ala Cys Leu Val Leu Val Ile Val Leu Leu Tyr Phe Ser Leu
                245                   250                   255
Trp Lys Gly Val Lys Thr Ser Gly Lys Val Val Trp Ile Thr Ala
                260                   265                   270
Thr Met Pro Tyr Val Val Leu Thr Ala Leu Leu Leu Arg Gly Val
                275                   280                   285
Thr Leu Pro Gly Ala Ile Asp Gly Ile Arg Ala Tyr Leu Ser Val
                290                   295                   300
Asp Phe Tyr Arg Leu Cys Glu Ala Ser Val Trp Ile Asp Ala Ala
                305                   310                   315
Thr Gln Val Cys Phe Ser Leu Gly Val Gly Phe Gly Val Leu Ile
                320                   325                   330
```

```
Ala Phe Ser Ser Tyr Asn Lys Phe Thr Asn Asn Cys Tyr Arg Asp
              335             340             345
Ala Ile Val Thr Thr Ser Ile Asn Ser Leu Thr Ser Phe Ser Ser
              350             355             360
Gly Phe Val Val Phe Ser Phe Leu Gly Tyr Met Ala Gln Lys His
              365             370             375
Ser Val Pro Ile Gly Asp Val Ala Lys Asp Gly Glu Pro Leu Leu
              380             385             390
Leu His Leu Gly Leu Leu Arg Val Ala Pro Ala Pro Ser Ser Gln
              395             400             405
Gln Thr Ser Ala Trp Ala Val Val Phe Phe Ile Met Leu Leu Thr
              410             415             420
Leu Gly Ile Asp Ser Ala Met Gly Gly Met Glu Ser Val Ile Thr
              425             430             435
Gly Leu Ile Asp Glu Phe Gln Leu Leu His Arg His Arg Glu Leu
              440             445             450
Phe Thr Leu Phe Ile Val Leu Ala Thr Phe Leu Leu Ser Leu Phe
              455             460             465
Cys Val Thr Asn Gly Gly Ile Tyr Val Phe Thr Leu Leu Asp His
              470             475             480
Phe Ala Ala Gly Thr Leu Met Asp Thr Asp Ala Thr Ser Ser Pro
              485             490             495
Ser Gly Val Gly Gln Phe Ser Asp Asp Ile Gln Gln Met Thr Gly
              500             505             510
Gln Arg Pro Ser Leu Tyr Trp Arg Leu Cys Trp Lys Leu Val Ser
              515             520             525
Pro Cys Phe Leu Leu Phe Val Val Val Val Ser Ile Val Thr Phe
              530             535             540
Arg Pro Pro His Tyr Gly Ala Tyr Ile Phe Pro Asp Trp Ala Asn
              545             550             555
Ala Leu Gly Trp Val Ile Ala Thr Xaa Ser Met Ala Met Val Pro
              560             565             570
Ile Tyr Ala Ala Tyr Lys Phe Cys Ser Leu Pro Gly Ser Phe Arg
              575             580             585
Glu Lys Leu Ala Tyr Ala Ile Ala Pro Glu Lys Asp Arg Glu Leu
              590             595             600
Val Asp Arg Gly Glu Val Arg Gln Phe Thr Val Arg Ser Arg Ser
              605             610             615
Leu Leu Gly Leu Ser Arg Gly Asn Ser Glu
              620             625
```

<210> 6
<211> 592
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475553CD1

<400> 6
```
Met Glu Lys Ala Arg Pro Leu Trp Ala Asn Ser Leu Gln Phe Val
1               5               10              15
Phe Ala Cys Ile Ser Tyr Ala Val Gly Leu Gly Asn Val Trp Arg
                20              25              30
Phe Pro Tyr Leu Cys Gln Met Tyr Gly Gly Gly Ser Phe Leu Val
                35              40              45
Pro Tyr Ile Ile Met Leu Ile Val Glu Gly Met Pro Leu Leu Tyr
                50              55              60
Leu Glu Leu Ala Val Gly Gln Arg Met Arg Gln Gly Ser Ile Gly
                65              70              75
Ala Trp Arg Thr Ile Ser Pro Tyr Leu Ser Gly Val Gly Val Ala
                80              85              90
Ser Val Val Val Ser Phe Phe Leu Ser Met Tyr Tyr Asn Val Ile
                95              100             105
Asn Ala Trp Ala Phe Trp Tyr Leu Phe His Ser Phe Gln Asp Pro
                110             115             120
```

```
Leu Pro Trp Ser Val Cys Pro Leu Asn Gly Asn His Thr Gly Tyr
            125             130              135
Asp Glu Glu Cys Glu Lys Ala Ser Ser Thr Gln Tyr Phe Trp Tyr
            140             145              150
Arg Lys Thr Leu Asn Ile Ser Pro Ser Leu Gln Glu Asn Gly Gly
            155             160              165
Val Gln Trp Glu Pro Ala Leu Cys Leu Leu Leu Ala Trp Leu Val
            170             175              180
Val Tyr Leu Cys Ile Leu Arg Gly Thr Glu Ser Thr Gly Lys Val
            185             190              195
Val Tyr Phe Thr Ala Ser Leu Pro Tyr Cys Val Leu Ile Ile Tyr
            200             205              210
Leu Ile Arg Gly Leu Thr Leu His Gly Ala Thr Asn Gly Leu Met
            215             220              225
Tyr Met Phe Thr Pro Lys Ile Glu Gln Leu Ala Asn Pro Lys Ala
            230             235              240
Trp Ile Asn Ala Ala Thr Gln Ile Phe Phe Ser Leu Gly Leu Gly
            245             250              255
Phe Gly Ser Leu Ile Ala Phe Ala Ser Tyr Asn Glu Pro Ser Asn
            260             265              270
Asn Cys Gln Lys His Ala Ile Ile Val Ser Leu Ile Asn Ser Phe
            275             280              285
Thr Ser Ile Phe Ala Ser Ile Val Thr Phe Ser Ile Tyr Gly Phe
            290             295              300
Lys Ala Thr Phe Asn Tyr Glu Asn Cys Leu Lys Lys Val Ser Leu
            305             310              315
Leu Leu Thr Asn Thr Phe Asp Leu Glu Asp Gly Phe Leu Thr Ala
            320             325              330
Ser Asn Leu Glu Gln Val Lys Gly Tyr Leu Ala Ser Ala Tyr Pro
            335             340              345
Ser Lys Tyr Ser Glu Met Phe Pro Gln Ile Lys Asn Cys Ser Leu
            350             355              360
Glu Ser Glu Leu Asp Thr Ala Val Gln Gly Thr Gly Leu Ala Phe
            365             370              375
Ile Val Tyr Thr Glu Ala Ile Lys Asn Met Glu Val Ser Gln Leu
            380             385              390
Trp Ser Val Leu Tyr Phe Phe Met Leu Leu Met Leu Gly Ile Gly
            395             400              405
Ser Met Leu Gly Asn Thr Ala Ala Ile Leu Thr Pro Leu Thr Asp
            410             415              420
Ser Lys Ile Ile Ser Ser His Leu Pro Lys Glu Ala Ile Ser Gly
            425             430              435
Leu Val Cys Leu Val Asn Cys Ala Ile Gly Met Val Phe Thr Met
            440             445              450
Glu Ala Gly Asn Tyr Trp Phe Asp Ile Phe Asn Asp Tyr Ala Ala
            455             460              465
Thr Leu Ser Leu Leu Leu Ile Val Leu Val Glu Thr Ile Ala Val
            470             475              480
Cys Tyr Val Tyr Gly Leu Arg Arg Phe Glu Ser Asp Leu Lys Ala
            485             490              495
Met Thr Gly Arg Ala Val Ser Trp Tyr Trp Lys Val Met Trp Ala
            500             505              510
Gly Val Ser Pro Leu Leu Ile Val Ser Leu Phe Val Phe Tyr Leu
            515             520              525
Ser Asp Tyr Ile Leu Thr Gly Thr Leu Lys Tyr Gln Ala Trp Asp
            530             535              540
Ala Ser Gln Gly Gln Leu Val Thr Lys Asp Tyr Pro Ala Tyr Ala
            545             550              555
Leu Ala Val Ile Gly Leu Leu Val Ala Ser Ser Thr Met Cys Ile
            560             565              570
Pro Leu Ala Ala Leu Gly Thr Phe Val Gln Arg Arg Leu Lys Arg
            575             580              585
Gly Asp Ala Asp Pro Val Ala
            590
```

<210> 7
<211> 2168

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7617689CB1

<400> 7
```
ccacgcgtcc gtaaggtggg atggatagca gggtctcagg cacaaccagt aatggagaga 60
caaaaccagt gtatccagtc atggaaaaga aggaggaaga tggcaccctg gagcggggc 120
actggaacaa caagatggag tttgtgctgt cagtggctgg ggagatcatt ggcttaggca 180
acgtctggag gtttccctat ctctgctaca aaaatggggg aggtgccttc ttcatcccct 240
acctcgtctt cctctttacc tgtggcattc ctgtcttcct tctggagaca gcactaggcc 300
agtacactag ccagggaggc gtcacagcct ggaggaagat ctgccccatc tttgagggca 360
ttggctatgc ctcccagatg atcgtcatcc tcctcaacgt ctactacatc attgtgttgg 420
cctgggccct gttctacctc ttcagcagct tcaccatcga cctgccctgg ggcggctgct 480
accatgagtg gaacacagaa cactgtatgg agttccagaa gaccaacggc tccctgaatg 540
gtacctctga gaatgccacc tctcctgtca tcgagttctg ggagcggcgg gtcttgaaga 600
tctctgatgg gatccagcac ctgggggccc tgcgctggga gctggctctg tgcctcctgc 660
tggcctgggt catctgctac ttctgcatct ggaaggggt gaagtccaca ggcaaggtgg 720
tgtacttcac ggccacattt ccttacctca tgctggtggt cctgttaatt cgaggggtga 780
cgttgcctgg ggcagcccaa ggaattcagt tttacctgta cccaaacctc acgcgtctgt 840
gggatcccca ggtgtggatg gatgcaggca cccagatatt cttctccttc gccatctgtc 900
ttgggtgcct ggcagctaca acaagtacca caacaactgc tacagggact 960
gcatcgccct ctgcttcctc aacagcggca ccagctttgt ggccggcttt gccatcttct 1020
ccatcctggg cttcatgtct caggagcagg gggtgcccat ttctgaggtg gccgagtcag 1080
gccctggcct ggctttcatc gcttacccgc gggctgtggt gatgctgccc ttctctcctc 1140
tctgggcctg ctgtttcttc ttcatggtcg ttctcctggg actggatagc cagtttgtgt 1200
gtgtagaaag cctggtgaca gcgctggtgg acatgtaccc tcacgtgttc cgcaagaaga 1260
accggaggga agtcctcatc cttggagtat ctgtcgtctc cttccttgtg gggctgatca 1320
tgctcacaga gggcggaaatg tacgtgttcc agctctttga ctactatgca gccagtggca 1380
tgtgcctcct gttcgtggcc atcttcgagt ccctctgtgt ggcttgggtt tacggagcca 1440
agcgcttcta cgacaacatc gaagacatga ttgggtacag gccatggcct cttatcaaat 1500
actgttggct cttcctcaca ccagctgtgt gcacagccac ctttctcttc tccctgataa 1560
agtacactcc gctgacctac aacaagaagt acacgtaccc gtggtggggc gatgccctgg 1620
gctggctcct ggctctgtcc tccatggtct gcattcctgc ctggagcctc tacagactcg 1680
gaaccctcaa gggccccttc agagagagaa tccgtcagct catgtgccca gccgaggacc 1740
tgcccagcg gaacccagca ggaccctcgg ctcccgccac ccccaggacc tcactgctca 1800
gactcacaga gctagagtct cactgctagg gggcaggccc ttggatggtg cctgtgtgcc 1860
tggccttggg gatggctgtg gagggaacgt ggcagaagca gccccatgtg cttccctgcc 1920
cccgacctgg agtggataag acaagagggg tattttggag tccacctgct gagctggagg 1980
cctcccactg caacttttca gctcaggggt tgttgaacag atgtgaaagg ccagtgccaa 2040
gagtgtccct ctgagaccct tgggaagctg ggtgggggct ggtaggtggg gcgagacttg 2100
ctggcttcgg gccctctcat ccttcattcc attaaatcca cattcttccc gctgaaaaaa 2160
aaaaaaaa                                                         2168
```

<210> 8
<211> 2709
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6881669CB1

<400> 8
```
catcctcggt ctttacttcc tatcaagtaa cattgtttg aaaaatagag ttaacacatt 60
tgccataagg gagtttttt ttttttttta aattacttcg aatactctcc aaatgcccaa 120
aaatagcaag gtggtaaaaa gagaattaga tgatgatgtt actgagtctg tcaaagacct 180
tctttccaat gaagacgcag ctgatgatgc ttttaagaca agtgaactaa ttgttgatgg 240
ccaggaagag aaagatacag atgttgaaga aggatctgaa gtcgaagatg aaagaccagc 300
ttggaacagt aaactacaat acatcctggc ccaagttgga ttttctgtag gtttaggaaa 360
tgtgtggcga tttccatacc tatgtcagaa gaatgggggc ggtgcatatc ttttaccata 420
tttaatacta cttatggtaa taggtattcc ccttttttc ttggaactct ctgtgggtca 480
aagaattcgg cgaggcagca ttggtgtatg gaattacata agccctaaac tgggcgggat 540
tggatttgca agttgtgtag tgtgctattt tgtagctctc tactacaacg tcatcattgg 600
ctggagtttg ttttatttttt ctcagtcttt tcagcaaccc ctgccttggg atcagtgtcc 660
```

```
tttggtgaaa aatgcttcac acacttttgt agaaccagaa tgtgaacaaa gttctgccac 720
cacctattac tggtacaggg aagcactgaa tatttcaagt tccatttctg aaagtggggg 780
cttaaactgg aagatgacca tctgcttgtt ggctgcctgg gtcatggttt gcttggctat 840
gatcaaaggc attcagtctt ctggaaaaat catatatttt agttctctgt ttccatatgt 900
ggtacttatt tgcttcctca tcagagcatt ccttttaaat ggttcaattg atggcattcg 960
ccacatgttt acccctaagc ttgaaataat gctggagccc aaggtctgga gagaagctgc 1020
tactcaagtg ttctttgcct taggtctggg atttggtggt gtcattgcct tttcaagcta 1080
caacaagaga gacaacaact gccactttga tgctgtcctg gtgtccttca tcattttttt 1140
cacttctgtc ctggcaacat tggtggtgtt tgcagttctg ggcttcaaag caaatgtcat 1200
aaatgagaaa tgcattacac aaaattcaga gacgatcatg aaatttttga aaatggggaa 1260
cattagtcag gatattattc cccatcatat caaccttttca actgttactg cagaagatta 1320
tcatttagtt tatgacatca ttcaaaaagt gaaagaagaa gagtttcctg ctcttcatct 1380
caattcctgt aaaattgaag aagagctaaa taaagctgtt caggggaccg gcttagcttt 1440
tattgccttt acagaagcga tgacacattt tcctgcatct cccttctggt cagtgatgt 1500
tttcctcatg ctggtcaatc taggcctttgg cagtatgttt ggaaccattg aagggattgt 1560
cacgcctatt gtggacactt tcaaagtgag gaaagaaatt cttactgtta tctgttgtct 1620
tctggcattt tgtattggcc tgatatttgt gcaacgctct ggaaattact ttgttacaat 1680
gtttgatgat tattctgcta cactgcctct gctaattgta gtcattttgg agaatattgc 1740
tgtatgcttt gtttatggca tagataagtt tatggaagac ctaaaagata tgctgggctt 1800
tgctcccagc agatattact actatatgtg gaaatatatt tctcctctaa tgctattatc 1860
attgctaata gctagtgttg tgaatatggg attaagtcct cctggctata acgcatggat 1920
tgaagataag gcatctgaag aattctgag ctatccaaca tggggactgg ttgtttgtgt 1980
ctctcggtt gtcttgcaa tactcccagt ccctgtagtt ttcattgttc gtcgcttcaa 2040
ccttatagat gatagttctg gtaatttagc atctgtgacc tataagagag gaagggtcct 2100
gaaagagcct gtgaacttag agggcgatga tacaagcctc attcacggaa aaataccgag 2160
cgagatgcca tctccaaatt ttggtaaaaa tatttatcga aaacagagtg gatccccaac 2220
tctggatact gctcccaatg gacggtatgg aataggtac ttgatggcag atattatgcc 2280
agatatgcca gaatctgatt tgtagctggg ggaaaagtca gtgggtttta tttggttcat 2340
ttttaccaat gaacattggc cctagtagga gaagcattag gcttcactta tcagagggca 2400
atctcaggtg ttccgtggct gtgatctttta atcctaacag tatatgtcag ttcaacttga 2460
gcattctttt ggattcttg gtttacattt gtgcagaaag gattgcagac aaatcttagg 2520
agggctgagg tacatgtttg ccaggatttt tttttaagta cctttggtgt atttttcaaat 2580
atttctatct cttaaaaaaa tggtattacc tcagtttcta ataatttctg ggtttagtag 2640
tgttgacaat taaaaatggt atacattaaa atttataagt tgccttcagg gtaacttcca 2700
gtgtcacac                                                             2709
```

<210> 9
<211> 2958
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2742486CB1

<400> 9

```
caggtccaag agattcttct atctcagcct cctgagtagc tgagattaca gtgctttat 60
tggaccaacg ctgggtggat ttctgtatga gaaaattggt tttgaatggg cagcagctat 120
acaaggtcta tgggctctga taagttgatt agccatgggc ttgttttttc tccttgaata 180
ctccagtaga taaaacaagc ccatggctaa tccacttatc agagcccata gaccttgtat 240
agctgctgcc cattcaaaac caattttctc atacagaaat ccacccagcg ttggtcccat 300
aaaagcacca attgaccaca ttgcactaaa aagacctgat acaagtccca atgtacttaa 360
tccctcttca aacccatttt catgtgcaca actgaagatt tccgggaaag ttggaattat 420
actcattcca gcagagaggc cacttacaac taatatcagc accagcagcc agagctgact 480
tttaatatgc aagattgcgg ggacccagag tcctggaagg ctttgggga g ctccgggccg 540
ggcggatcgc tgcctgcagg gagtcgggga tgccaggttc cagctgagca gcggccgccc 600
gccagagtgc cagtggctcc ttggaggtcg agtccaagga cgtggcttga agccgggagc 660
tggggcgccg gagtccacgc accggggatg gaggcgctgg gtgacctgga gggaccacgc 720
gcaccaggag gtgatgatcc tgcaggaagt gcaggagaga cccccgggtg gctttcgaga 780
gaacaggttt ttgtactgat atcggcagct tcggtgaact taggttccat gatgtgctat 840
tctatacttg gaccgttttt ccccaaagag gctgaaaaga agggagccag caatacaatt 900
atcggtatga tctttggatg ttttgctttg ttcgagttgc tggcatcctt ggtatttgga 960
aactatctgg tacatattgg agcaaaattt atgtttgtag caggaatgtt tgtctcagga 1020
ggagttacaa ttctctttgg tgtattggac cgagttccag atgggccagt atttattgct 1080
atgtgttttc tagtgagagt aatggatgca gttagctttg ctgcagcaat gactgcatct 1140
tcttctatcc tggcaaaggc ttttccaaat aacgtggcta cggtattggg aagtcttgag 1200
acttttctg gactgggggct aatactaggt cctcctgtag gtggcttttt gtatcaatcc 1260
```

```
tttggctatg aagtgccttt tattgttctg ggatgcgtcg ttttgctgat ggtaccactc 1320
aatatgtata ttttacccaa ttacgagtct gatccaggtg aacactcatt ctggaaactg 1380
atcgctttac ccaaagttgg ccttatagcc ttcgtcatca actcactcag ctcgtgtttt 1440
ggcttcctcg atcctactct gtctctcttt gttttggaga agttcaattt accagctgga 1500
tatgtgggac tagtattcct gggtatggca ctgtcctatg ccatctcttc accactattt 1560
ggtctcctaa gtgataaaag gccacctcta aggaaatggc ttctggtgtt tggcaactta 1620
atcacagccg ggtgctacat gctcttaggg cctgtcccaa tcttgcatat taaaagtcag 1680
ctctggctgc tggtgctgat attagttgta agtggcctct ctgctggaat gagtataatt 1740
ccaactttcc cggaaattct cagttgtgca catgaaaatg ggtttgaaga gggattaagt 1800
acattgggac ttgtatcagg tcttttttagt gcaatggggt caattggtgc ttttatggga 1860
ccaacgctgg gtggatttct gtatgagaaa attggttttg aatgggcagc agctatacaa 1920
ggtctatggg ctctgataag tggattagcc atgggcttgt tttatctact ggagtattca 1980
aggagaaaaa ggtctaaatc tcaaaacatc ctcagcacag aggaggaacg aactactctc 2040
ttgcctaatg aaacctagtc cgatggatcc tggattgata caaggttgag aaatgaatgc 2100
tcctggcctt aaacatcacc gtaggaaggg tttttaaaat tttacgcgca aaactccgtg 2160
gaccccgtgc cagtgtcttg gaagtgtcaa cgtgtttttg gatgatcctg tattgggctg 2220
tacttactgt gatactgaaa agctgtcctg ctgaagcagc tatatttgaa atattaagta 2280
tgaaaggagt aattaaaaac aagcaaaaca aaacaagact tagttttttaa atgaccaaac 2340
ttgtccttaa agatgttgtt attaactcga gttagttctt atttcctctg tttattttt 2400
attctaagta cactgattct gtgaatgtac cttttttatt aacagggaaa gaaatgaatt 2460
aatttgatat gctctaaata cataaaggtg cttcaaaata tgtagaaaca ttactatgaa 2520
atcagttttt aaaagatata ctttctcttt gtcctgaggt ttttcggtct tgttcaaaag 2580
gaagaattct tgcctgccat acagaaactc tctagcactc cctgacctta agcttttcta 2640
aaaattctgt ttgtgtgaaa agtacaagaa taacaatact tacaacttcc atttttgtaa 2700
cctacgttca cttatgatct ggatttataa acattacttg gtataacgtt tttcatttcc 2760
tttaatgtct ctgttttttg gctctaccat ctgttttgtt tttgttttta tctatatctt 2820
ggtagatgta tttcatccct agagcaggtc agcctccttc ccctaatgcg aatgcttgtt 2880
ttgttaggga agggcttcct ccaacttcgt gtgaaattgt gatgttgaag tgaataaatg 2940
tctattgtgt aaaaaaaa                                               2958
```

```
<210> 10
<211> 2135
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71556695CB1

<400> 10
ggaaaagggg cttggctccc tctggaccct tcacccagcc cctgcaccct ggagcagagc 60
tgcaggcagt gtgaccgttg tgctgttatt ggcctaaaca catagcaaca ggaaaacaac 120
cccaaaaccc cagccttaag agcaagtcca acagtgcgcg actccccagg tcaggagaag 180
ccccgatgtg tgcagaatct gcagatgggg ctgctctcgg ccctgcctct ggctgtggaa 240
acaggcctcc aagtacccaa gaaagcaagg agatggccac caaggagaag ctgcagtgtc 300
tgaaagattt ccacaaggac atcctgaagc cctcaccagg gaagagccca ggcacggcggc 360
ctgaggacga ggctgaggga aaacctccgc agagggagaa gtggtctagc aagatcgact 420
ttgtgctctc tgtggctggc ggcttcgtgg gcttgggcaa cgtctggcgc ttcccgtacc 480
tctgctacaa gaatggtgga ggtgcgtttc tcataccgta ttttattttc ctgtttgggg 540
gcggcctgcc tgtgtttttc ttggagatca tcataggcca gtacacctct gaaggggggca 600
tcacctgctg ggaaaagatc tgcccccttgt tctctggtat cggctatgcc tccgttgtaa 660
ttgtgtccct cctgaatgtc tactacatcg tcatcctggc ctgggccaca tactacctgt 720
tccagtcctt ccagaaggag ctgccctggg cacactgcaa ccacagctgg aacacacctc 780
actgcatgga ggacaccatg cgcaagaaca agagtgtctg gatcaccatc agctccacca 840
acttcacctc ccctgtcatc gagttctggg agcgcaacgt gctgagcttg tcccctggaa 900
tcgaccaccc aggctctctg aaatgggacc tcgctctctg ccttcttta gtctggctag 960
tgtgtttctt ctgcatctgg aagggcgtca ggtccactgg gaaggtcgtc tacttcacag 1020
ccacttttcc attcgccatg ctcctggtgc tgctggtccg agggctgacg ctgccgggcg 1080
cgggcgcagg catcaagttc tatctgtatc ctgacatcac ccgccttgag gacccacagg 1140
tgtggattga cgctgggact cagatattct tctcttatgc catctgcctg ggggctatga 1200
cctcgctggg gagctacaac aagtacaagt ataactcgta cagggactgt atgctgctgg 1260
gatgcctgaa cagtggtacc agttttgtgt ctggcttcgc aatttttttcc atcctgggct 1320
tcatggcaca agagcaaggg gtggacattg ctgatgtggc tgagtcaggt cctggcctga 1380
ccttcattgc ctacccaaaa gctgtgacaa tgatgccgct gcccacattt tggtccattc 1440
tttttttttat tatgcttctc ttgcttggac tggatagcca gtttgttgaa gttgaaggac 1500
agatcacatc cttggttgat ctttacccat ccttcctaag gaagggttat cgtcgggaaa 1560
tcttcatcgc cttcgtgtgt agcatcagct acctgctggg gctgacgatg gtgacggagg 1620
```

```
gtggcatgta tgtgtttcag ctctttgact actatgcagc tagcggtgta tgcctttttgt 1680
gggttgcatt ctttgaatgt tttgttattg cctggatata tggaggtgat aacctttatg 1740
atggtattga ggacatgatt ggctatcggc ccgggccctg gatgaagtac agctgggctg 1800
tgatcactcc agttctctgt gttggatgtt tcatcttctc gctcgtcaag tacgtacccc 1860
tgacctacaa caaaacatac gtgtacccca actgggccat tgggctgggc tggagcctgg 1920
cccttttcctc catgctctgc gttcccttgg tcatcgtcat ccgcctctgc cagactgagg 1980
ggccgttcct tgtgagagtc aagtacctgc tgaccccaag ggaacccaac cgctgggctg 2040
tggagcgcga gggagccaca ccttacaact ctcgcaccgt catgaacggc gctctcgtga 2100
aaccgaccca catcattgtg gagaccatga tgtga              2135
```

<210> 11
<211> 1997
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472800CB1

<220>
<221> unsure
<222> 1810
<223> a, t, c, g, or other

<400> 11
```
gagcgggagg ggaggcttcg cggaacgctc tcggcgccag gactcgcgtg caaagcccag 60
gcccgggcgg ccagaccaag agggaagaag cacagaattc ctcaactccc agtgtgccca 120
tgagtaagag caaatgctcc gtgggactca tgtcttccgt ggtggccccg gctaaggagc 180
ccaatgccgt gggcccgaag gaggtggagc tcatccttgt caaggagcag aacggagtgc 240
agctcaccag ctccaccctc accaacccgc ggcagagccc cgtggaggcc caggatcggg 300
agacctgggg caagaagatc gactttctcc tgtccgtcat tggctttgct gtggacctgg 360
ccaacgtctg gcggttcccc tacctgtgct acaaaaatgg tggcggtgcc ttcctggtcc 420
cctacctgct cttcatggtc attgctggga tgccactttt ctacatggag ctggccctcg 480
gccagttcaa cagggaaggg gccgctggtg tctggaagat ctgccccata ctgaaaggtg 540
tgggcttcac ggtcatcctc atctcactgt atgtcggctt cttctacaac gtcatcatcg 600
cctgggcgct gcactatctc ttctcctcct tcaccacgga gctccctgg atccactgca 660
acaactcctg gaacagcccc aactgctcgg atgcccatcc tggtgactcc agtggagaca 720
gctcgggcct caacgacact tttgggacca cacctgctgc cgagtacttt gaacgtggcg 780
tgctgcacct ccaccagagc catggcatcg acgacctggg gcctccgcgg tggcagctca 840
cagcctgcct ggtgctggtc atcgtgctgc tctacttcag cctctggaag ggcgtgaaga 900
cctcagggaa ggtggtatgg atcacagcca ccatgccata cgtggtcctc actgccctgc 960
tcctgcgtgg ggtcaccctc cctggagcca tagacggcat cagagcatac ctgagcgttg 1020
acttctaccg gctctgcgag gcgtcggttt ggattgacgc ggccacccag gtgtgcttct 1080
ccctgggcgt ggggttcggg gtgctgatcg ccttctccag ctacaacaag ttcaccaaca 1140
actgctcagg ggacgcgatt gtcaccacct ccatcaactc cctgacgagc ttctcctccg 1200
gcttcgtcgt cttctccttc ctggggtaca tggcacagaa gcacagtgtg cccatcgggg 1260
acgtggccaa ggacggtgag cccctcctgc tgcacctggg cctgctccgt gtagcaccag 1320
cgccgagctc tcagcaaacc tcagcctggg ccgtggtctt cttcatcatg ctgctcaccc 1380
tgggtatcga cagcgccatg ggtggtatgg agtcagtgat caccgggctc atcgatgagt 1440
tccagctgct gcacagacac cgtgagctct tcacgctctt catcgtcctg gcgaccttcc 1500
tcctgtccct gttctgcgtc accaacggtg gcatctacgt cttcacgctc ctggaccatt 1560
ttgcagccgg cacgttgatg gacactgatg ccacctcttc tccctcaggt gttgggcagt 1620
tcagcgacga catccagcag atgaccgggc agcggcccag cctgtactgg cggctgtgct 1680
ggaagctggt cagcccctgc tttctcctgt tcgtggtcgt ggtcagcatt gtgaccttca 1740
gacccccca ctacggagcc tacatcttcc ccgactgggc caacgcgctg ggctgggtca 1800
tcgccacatn ctccatggcc atggtgccca tctatgcggc ctacaagttc tgcagcctgc 1860
ctgggtcctt tcgagagaaa ctggcctacg ccattgcacc cgagaaggac cgtgagctgg 1920
tggacagagg ggaggtgcgc cagttcacgg tgaggtcgag gtccctgctg ggcctctctc 1980
gggggaattc agaatga              1997
```

<210> 12
<211> 2774
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 7475553CB1

<220>
<221> unsure
<222> 2666
<223> a, t, c, g, or other

<400> 12
```
cgagccggag ccgagcgcgc cgaggccggg gccatggaga aagcgcggcc gctgtgggcc   60
aactcgctac agttcgtgtt cgcctgcatc tcgtacgccg tgggcctggg caacgtgtgg  120
cgattcccgt acctgtgcca gatgtacggc ggaggtagtt tcctggtccc ctacatcatc  180
atgcttatcg tggagggaat gccgctcttg tacctggaac tggctgtggg gcagcgcatg  240
cggcagggca gcatcggcgc ctggaggacc atcagcccgt acctcagtgg tgtcggggtc  300
gccagcgtgg tggtctcttt cttcctctcc atgtactaca acgtcatcaa cgcctgggcc  360
ttctggtacc tcttccactc cttccaggat cccctgccgt ggtctgtctg cccactgaat  420
ggtaaccaca cgggctacga tgaggagtgc gagaaggcgt cctccacaca gtacttctgg  480
tacaggaaaa ccctcaatat ctcgccgtcc ctccaggaga acggggggtgt gcagtgggag  540
ccggcgctgt gcctcctcct ggcctggctg gtggtgtacc tgtgcatcct gcgtggcacc  600
gagtccactg gcaaggtggt gtatttcacg gcgtcactgc cctattgcgt gctcatcatc  660
tacctcatca ggggcctcac gctccacgga gccaccaatg gcctcatgta catgttcact  720
cccaagatag agcagctggc caaccccaag gcctggatca atgcagccac ccagatcttc  780
ttctcacttg gcctgggctt cggcagcctg atcgccttcg ccagctacaa tgagccatcc  840
aacaactgcc agaagcacgc catcatcgtg tccctcatca acagcttcac ctccatattt  900
gccagcattg tcaccttctc catctatggc ttcaaggcca ccttcaatta tgaaaactgc  960
ttgaagaagg tgagtctgct gctgaccaac acttttgacc ttgaagatgg cttttttgaca 1020
gccagcaacc tggagcaggt gaagggctac ctcgcatctg cctacccaag caaatacagc 1080
gagatgttcc cgcaaatcaa aaactgcagc ttggaatcgg agctagacac ggccgtccag 1140
ggcactggcc tggcattcat cgtctacaca gaggccatta aaaacatgga ggtgtcccag 1200
ctgtggtcgg tgctctactt cttcatgctg ctgatgctgg gcattgggag catgctgggg 1260
aacacagcgg ccatcctcac ccctctgaca gacagcaaga tcatctccag ccacctgccc 1320
aaggaggcca tctcaggtct ggtgtgcctt gtcaactgtg ccattggcat ggtgttcacg 1380
atggaggctg ggaactactg gtttgacata ttcaacgact acgcggccac actgtccctg 1440
ctgctcatcg tgctggtgga gacgattgcc gtgtgctacg tgtacgggct gaggagattt 1500
gaaagtgacc ttaaggccat gaccggccga gctgtgagct ggtactggaa ggtgatgtgg 1560
gctggcgtaa gcccactgct gattgtcagc ctctttgtct tctacctgag cgactacatc 1620
ctcacgggga ccctgaagta tcaagcctgg gacgcctccc agggccagct cgtgaccaaa 1680
gattacccgg cctatgcact ggctgtcatc gggctgcttg tggcctcctc caccatgtgc 1740
atcccctgg cggccctggg gacttttgtt cagcgtcgcc tcaagagggg agacgcagac 1800
cccgtggcct gagatgtggg cttcccagcc gctcacggtt ttacagatac tatttacagg 1860
cggaaactcc tcggctgctt tttcaaatgc ttaagccagg agtgctcagc ccatcaactt 1920
cctgagtgtc taaagaagat gaggaaggtg tgcaggaaga aaactccctt gggagaacgc 1980
acacctcccc gtggtggctg ttcctccctg tcacctgcct cctcatcatg gaaggggtg 2040
ggctatgaaa gccggtctca aagataactg catccttcat tccaggaaag ccctagaatt 2100
agggcacatt gcaaactgaa atatgactat aattcttatg ggaccaaatt taagcaattt 2160
ttgttttgg ctgaagagac accaaaatat tagaggacaa atatttttag atccatttaa 2220
ggagtttga agtgcctaag atgacctatt tgtcagtggt gcaaaattaa ttctcttctt 2280
ttttgagttg tagtgaatat gcaatttctt tgttcccctt ccacccttta aatcttagga 2340
tgacaagtca taaagaaaga agatctttgt ctgggacccc caaagggatc ctttctctaa 2400
ggtctctgac agtgggtcca ggaccagacc tctctacaaa aaattgcccc aactacagtt 2460
tgcaaccccca aaccacatta gaagtctgtg cagacatccc tccgtggtgt gtgtcttggt 2520
gcattggaaa aggagtcagg agccactgtg aggtgagaat gaaagtggat ctcagctggg 2580
cacggtggct cacgcctgta atcctagcac cttgggggtc aaggtgggtg gatcacttga 2640
ggtcaggagt ttgaaggcag cctggncaaa atggcgaaaa cccatttcta ctaaaaatcc 2700
aacaaattag ctggcattgg gggagggttg gaggaatcga attaaagtta tggaaagggg 2760
gacccagggg gggg                                                    2774
```